ΠΑΝΤΕΙΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES

SCHOOL OF POLITICAL SCIENCE

DEPARTMENT OF SOCIAL POLICY

Estimating Poverty and Unemployment using Small Area Estimation Methods

DOCTORAL DISSERTATION

Eleni Malapani

Athens, 2021

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Scholarships programme for post-graduate studies - 2nd Study Cycle" (MIS-5003404), implemented by the State Scholarships Foundation (IKY).

Ευρωπαϊκή Ένωση
European Social Fund

Operational Programme
Human Resources Development, Education and Lifelong Learning
Co-financed by Greece and the European Union

ΕΣΠΑ
2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

ΠΑΝΤΕΙΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES

SCHOOL OF POLITICAL SCIENCE

DEPARTMENT OF SOCIAL POLICY

Estimating Poverty and Unemployment using Small Area Estimation Methods

DOCTORAL DISSERTATION

Eleni Malapani

Athens, 2021

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Scholarships programme for post-graduate studies - 2nd Study Cycle" (MIS-5003404), implemented by the State Scholarships Foundation (IKY).

Ευρωπαϊκή Ένωση
European Social Fund

Operational Programme
Human Resources Development, Education and Lifelong Learning
Co-financed by Greece and the European Union

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

Supervising committee

Catherine Michalopoulou, Professor, Panteion University (Supervisor)

Stefanos Giakoumatos, Professor, University of the Peloponnese

Clive Richardson, Professor Emeritus, Panteion University

To John and Maximos

Abbreviations

AROP: At Risk of Poverty

AROPE: At Risk of Poverty or social Exclusion

BLUP: Best Linear Unbiased Predictor

CV: Coefficient of Variation

EB: Empirical Bayes

EBLUP: Empirical Best Linear Unbiased Predictor

ELSTAT: Hellenic Statistical Authority

ESSnet: European Statistical System network

EURAREA: Enhancing Small Area Estimation Techniques to meet European needs

FGT poverty measures: Foster, Greer and Thorbecke poverty measures

F-H: Fay and Herriot

HB: Hierarchical Bayes

H-T: Horvitz-Thompson

ILO: International Labour Organization

LAUs: Local Administrative Units

LFS: Labour Force Survey

ML: Maximum Likelihood

MSE: Mean Square Error

NSI: National Statistical Institute

NUTS: Nomenclature of Territorial Units for Statistics

OECD: Organization for Economic Co-operation and Development

ONS: Office for National Statistics (United Kingdom)

REML: Residual Maximum Likelihood

SAE: Small Area Estimation

UNECE: United Nations Economic Commission for Europe

Contents

Tables

Figures

Abstract

Fighting poverty and unemployment are two of the goals of the European Commission, which is clearly emphasized in the “Europe 2020” strategy. The surveys carried out for these social characteristics provide very reliable information at national level but due to the design and the sample size they cannot give reliable estimates in smaller geographical areas of the country, such as the prefectures (NUTS 3) and municipalities. The key solution in order to produce reliable estimates in small areas is to combine data from administrative, Census and survey sources using Small Area Estimation methods (SAE). In Greece the lowest geographical level for which estimates of poverty and unemployment are given is the NUTS 2 level (*perifereies*). The main goal of this thesis is to develop and produce reliable estimates for the poverty and unemployment in Greece at a lower level of spatial aggregation than the one used so far, that is at the level of sub regions-NUTS 3 (*Nomoi*), using SAE methods. The target parameters to be estimated are headcount ratio, poverty gap index and unemployment rate of the Greek population at two different times, in 2009 (shortly before the start of the Greek financial crisis) and 2013 (during the crisis)). To achieve the above objectives the EBLUP estimator based on the Fay and Herriot (F-H) small area model was adopted, combining survey data from the EU-SILC 2009 and 2013 with auxiliary data derived from the 2001 and 2011 national Greek Census, respectively. Specifically, 19 auxiliary variables from the Greek Census of 2001 and 32 auxiliary variables from the Greek Census of 2011 were considered and analyzed. In order to build the optimal small area model a variable selection process was performed for each of the target parameters. Then, different diagnostic tools applied to assess the fit and the performance of the selected small area models as well as to check the reliability of the results. These diagnostics showed that the final selected models provide a good fit to the data and produce reliable estimates. The research results were particularly encouraging as the application of small area estimation approaches achieved an overall significant efficiency gain both for estimating poverty and unemployment instead of direct estimators. Specifically, results showed a significant reduction in coefficient of variation (CV) of the EBLUP F-H estimators over direct estimators for most domains. The reduction tended to be greater for domains with small sample sizes. Also, the results of estimates of both poverty and unemployment showed significant differences in the map of Greece in 2009 and 2013. The present study contributes to the growing demand for estimates of social characteristics in small geographical areas by developing appropriate SAE models and giving estimates for poverty and unemployment in Greece for the first time at the level of sub-regions-NUTS 3 (*nomoi*). These estimates allow governments to formulate and target policies and thus allocate funds properly to small areas.

Keywords: SAE methods, Fay-Herriot model, EBLUP, poverty, unemployment, EU-SILC, Census

Εκτιμώντας τη φτώχεια και την ανεργία χρησιμοποιώντας μεθόδους εκτίμησης σε μικρές γεωγραφικές περιοχές

Ελένη Μαλαπάνη

Περίληψη

Τις τελευταίες δεκαετίες η καταπολέμηση της φτώχειας και της ανεργίας αποτελεί μία από τις κυριότερες προκλήσεις της Ευρώπης και όχι μόνο. Η Ευρωπαϊκή Επιτροπή στα πλαίσια της στρατηγικής «Ευρώπη 2020» έθεσε το παραπάνω πρόβλημα ως έναν από τους βασικούς στόχους της. Για την εκτίμηση χαρακτηριστικών όπως η φτώχεια και η ανεργία τα κράτη έχουν σχεδιάσει ειδικές έρευνες οι οποίες δίνουν μεν αξιόπιστες πληροφορίες σε εθνικό επίπεδο, αλλά λόγω σχεδιασμού και μεγέθους των δειγμάτων δεν μπορούν να δώσουν αντίστοιχες αξιόπιστες εκτιμήσεις σε μικρότερες γεωγραφικές περιοχές όπως π.χ. οι νομοί (NUTS 3) και οι δήμοι. Για τη διαχείριση και επίλυση του παραπάνω προβλήματος, δηλαδή την επίτευξη στατιστικών εκτιμήσεων κοινωνικών και οικονομικών δεικτών σε μικρές γεωγραφικές περιοχές, προτείνεται ο συνδυασμός των παραπάνω τύπων δεδομένων (ετήσιες έρευνες και απογραφικά δεδομένα). Ο συνδυασμός αυτός μπορεί να επιτευχθεί με τη χρήση προηγμένων στατιστικών μεθόδων που παράγουν εκτιμήσεις σε μικρές γεωγραφικές περιοχές και έχουν τη γενική ονομασία «Small Area Estimation» (SAE). Στην Ελλάδα το μικρότερο γεωγραφικό επίπεδο για το οποίο δίνονται εκτιμήσεις της φτώχειας και της ανεργίας είναι αυτό των περιφερειών. Ο κύριος στόχος αυτής της διατριβής είναι να αναπτύξει και να παράσχει αξιόπιστες εκτιμήσεις για τη φτώχεια και την ανεργία στην Ελλάδα σε μικρότερο γεωγραφικό επίπεδο από αυτό των περιφερειών, δηλαδή σε επίπεδο Νομών (NUTS 3) χρησιμοποιώντας τις μεθόδους SAE. Τα υπό εκτίμηση χαρακτηριστικά είναι το ποσοστό της φτώχειας, το χάσμα της φτώχειας καθώς και το ποσοστό της ανεργίας του ελληνικού πληθυσμού σε δύο διαφορετικές χρονικές στιγμές, το 2009 (λίγο πριν από την έναρξη της ελληνικής χρηματοπιστωτικής κρίσης) και το 2013 (κατά τη διάρκεια της κρίσης)). Για την επίτευξη των παραπάνω υιοθετήθηκε ο εκτιμητής EBLUP με βάση το μοντέλο Fay and Herriot, συνδυάζοντας δεδομένα από την έρευνα EU-SILC 2009 και 2013 με βοηθητικά δεδομένα από την εθνική απογραφή του 2001 και 2011, αντίστοιχα. Συγκεκριμένα εξετάσθηκαν και αναλύθηκαν 19 βοηθητικές μεταβλητές από την απογραφή του 2001 και 32 βοηθητικές μεταβλητές από την απογραφή του 2011. Προκειμένου να κατασκευαστεί το βέλτιστο μοντέλο μικρής περιοχής (small area model) για κάθε ένα από τα υπό εκτίμηση χαρακτηριστικά,

χρησιμοποιήθηκε μια διαδικασία τριών φάσεων για την επιλογή των τελικών βοηθητικών μεταβλητών. Έπειτα διάφοροι διαγνωστικοί έλεγχοι εφαρμόστηκαν με σκοπό την αξιολόγηση της καταλληλόλητας και απόδοσης των επιλεγμένων SAE μοντέλων καθώς και της αξιοπιστίας των αποτελεσμάτων. Τα αποτελέσματα αυτών των διαγνωστικών ελέγχων έδειξαν ότι τα επιλεγμένα μοντέλα παρέχουν καλή προσαρμογή στα δεδομένα καθώς και αξιόπιστες εκτιμήσεις. Τα αποτελέσματα της έρευνας ήταν ιδιαίτερα ενθαρρυντικά καθώς η εφαρμογή των μεθόδων SAE πέτυχε ένα στατιστικά σημαντικό συνολικό κέρδος απόδοσης τόσο για την εκτίμηση της φτώχειας όσο και της ανεργίας έναντι των άμεσων εκτιμητών. Συγκεκριμένα, τα αποτελέσματα έδειξαν μία στατιστικά σημαντική μείωση τόσο των τιμών του συντελεστή μεταβλητότητας (CV) όσο και των τιμών του μέσου τετραγωνικού σφάλματος (MSE) του EBLUP εκτιμητή με βάση το μοντέλο F-H έναντι των άμεσων εκτιμητών σχεδόν σε όλους τους νομούς. Η μείωση ήταν αισθητά μεγαλύτερη στους νομούς με μικρό μέγεθος δείγματος. Επίσης, τα αποτελέσματα των εκτιμήσεων τόσο της φτώχειας όσο και της ανεργίας έδειξαν σημαντικές διαφορές στο χάρτη της Ελλάδας τις χρονιές 2009 και 2013. Η παρούσα μελέτη συμβάλλει στην ολοένα και αυξανόμενη ζήτηση για εκτιμήσεις κοινωνικών χαρακτηριστικών σε μικρές γεωγραφικές περιοχές αναπτύσσοντας κατάλληλα SAE μοντέλα και δίνοντας εκτιμήσεις για τη φτώχεια και την ανεργία στην Ελλάδα για πρώτη φορά σε επίπεδο νομών. Οι εκτιμήσεις αυτές μπορούν να συμβάλουν στη διαμόρφωση και στόχευση πολιτικών για τη σωστή κατανομή των δημόσιων κονδυλίων σε μικρές γεωγραφικές περιοχές.

Λέξεις-κλειδιά: Μέθοδοι SAE, μοντέλο Fay-Herriot, φτώχεια, ανεργία, EU-SILC, απογραφή.

1. Introduction

Statistical information has always played a vital role in the social development of our society. More and more policy makers are demanding estimates for small areas, to measure a number for characteristics that determine poverty, unemployment and living conditions of households in order to make policy decisions (Rao and Molina, 2015). The National Statistical Authorities (in the case of Greece, ELSTAT) produce annual statistical estimates for these characteristics at national level.

These surveys provide very reliable information at national level but due to their design and the sample size they cannot give reliable estimates in smaller geographical areas of the country, such as the prefectures (NUTS 3) and municipalities (Ghosh and Rao, 1994). On the one hand Census data could provide a solution in this problem, but in most countries, Censuses are conducted only once a decade. Therefore, they do not cover the changes during the course of the decade. Furthermore, Census content is restricted. Most of the time they do not examine enough variables related to household living conditions, thus estimates for the characteristics of interest cannot be obtained. On the other hand, increasing the sample size would be prohibitively expensive. The key solution in order to produce reliable estimates in small areas is to combine data from administrative, Census and survey sources using Small Area Estimation methods (SAE) (Kordos, 2016).

As pointed out by Whitworth (2013) "SAE methodologies have become increasingly demanded, increasingly used and increasingly refined" as they have given insights that would not otherwise be possible. For instance, fighting poverty and unemployment are two of the goals of the European Commission, which is clearly emphasized in the "Europe 2020" strategy[1]. The sample size of the surveys carried out for these social characteristics allows for accurate estimates only at a very general level such as the whole country and regions. SAE methods can give estimates at a lower level of spatial aggregation so that policy makers and stakeholders can formulate and implement policies, and distribute resources in the small areas as well (Pratesi, 2016). In this context, many projects around the world have been carried out under the auspices

[1] In June 2010, the European Council adopted the Europe 2020 Strategy which is the EU's growth strategy for the current decade, aiming at developing in the EU a smart, sustainable and inclusive economy. In this context, the European Council adopted a social inclusion target, namely lifting at least 20 million people from the risk of poverty and exclusion by 2020 (European Commission, 2010).

of National Statistical Institutes using SAE methods, such as ESSnet, SAIPE, SAMPLE, AMELI, BIAS and the EURAREA project[2].

The term "SAE" is somewhat confusing, since it is the size of the sample in the area that causes estimation problems, and not the size of the area. SAE methods are used for producing estimates of population parameters for areas (domains) with small, or even zero, sample size. In those areas, direct estimators that rely only on domain-specific observations may lead to estimates with large sampling variability (Pfeffermann, 2002). When direct estimation is not possible, one has to rely upon "indirect" estimators.

> "Indirect" estimators "borrow strength" by using values of the variable of interest, y, from related areas and/or time periods and thus increase the effective sample size. These values are brought into the estimation process through a model (either implicit or explicit) that provides a link to related areas and/or time periods using the supplementary information related to y, such as recent Census database and current administrative records (Rao, 2003, p. 2).

Having only a small sample in each area, the only possible solution to the estimation problem is to borrow information from other related data sets: data from other 'similar' areas or from previous occasions. Since SAE methods depend on auxiliary data, their availability is a major issue. As Whitworth (2013, p. 4) pointed out " the availability of small area data has improved dramatically since the late 1990s yet many spatial variables of interest – income, fear of crime, health-related behaviours, and so the list goes on – remain impossible to access at small area geographies in many national contexts". The availability of good auxiliary data is crucial to the formation of small area models. The success of any model-based method depends on how good the auxiliary data are as predictors of the study variables (Rao, 2003; Whitworth, 2013).

Data for poverty measurement may come from sample surveys at national level, meso databases at regional level, local databases, registers and sample surveys conducted at regional and local levels. In the EU there are several cross-national comparative surveys on the study of poverty and social exclusion. However, these surveys either cover only a portion of the population or have a small sample size or contain only limited information on income and living conditions (Pratesi, 2016). From 2004 onwards, the EU's main source for micro-data on income and living conditions is

[2] Details about these projects are given in section 2.7.

the annual EU-SILC[3] (European Union Statistics on Income and Living Conditions) survey. In Greece, the minimum effective sample size of EU-SILC aims to provide accurate estimates at regional level (NUTS 2).

In addition, common sources of national labour market information are Labour Force Surveys (LFS) and other household sampling surveys, as well as population Censuses. However, they are unable to deliver direct estimates of unemployment with adequate precision for every local authority district because the sample size in many areas is insufficient (ONS, 2006). In Greece the lowest geographic areas for which the LFS publishes estimates are NUTS 2 areas.

Indicators of poverty and unemployment have an important territorial dimension associated with the need to take into account regional and local differences in the construction of the system of these indicators (SAMPLE, 2009). Many countries around the world decentralize decision-making, resources and responsibilities to lower levels of government, and as a result, regional and local governments have more opportunities to tackle poverty and unemployment. Thus, in order to ensure a good allocation of public resources, a system of indicators of poverty, unemployment and social exclusion at regional and local level is necessary. Therefore, small area estimation techniques for measuring poverty at local level are required that “borrow strength” across areas through linking models and auxiliary information such as Censuses and administrative data. Through using this information from other sources, it is possible to estimate distribution parameters with smaller variance than in the case of direct estimation (Molina and Rao, 2010).

In the present thesis SAE methods were used to estimate poverty and unemployment in Greece at a lower level of spatial aggregation than the one used so far, that is at the level of sub regions-NUTS 3 (*Nomoi*). A linear mixed model called Fay-Herriot was applied to produce estimates of the headcount ratio, the poverty gap and the unemployment rate of the Greek population at two different times, in 2009 (shortly before the start of the Greek financial crisis[4]) and 2013 (during the crisis)). The above work was achieved combining survey data from the EU-SILC 2009 and 2013 with auxiliary data derived from the 2001 and 2011 national Greek Census, respectively. Specifically, the thesis is organized as follows: Chapter 2 describes the

[3] Details of the EU-SILC survey are given in section 3.4.
[4] A full-blown financial crisis took place in Greece by the end of 2009 (Gibson et al., 2014)

basic theoretical framework for the small area estimation methods. Two broad types of SAE methods-Direct and Indirect- are presented. Basic direct (H-T, GREG and Modified direct) and indirect estimators (Synthetic, Composite, Area level and Unit level) are analyzed. Also, the problem of selecting appropriate auxiliary information is discussed as well as the assessment of the quality and plausibility of the estimates. At the end of Chapter 2, some of the most important SAE projects around the world are presented. Chapter 3 analyzes various issues related to the measurement of poverty and focuses on the application of SAE methods for estimating poverty in small geographical areas. Chapter 4 presents, analyzes, and interprets the results of the application of SAE methods for the estimation of poverty in Greece at NUTS 3 level at two different times, in 2009 and 2013, combining data from the EU-SILC survey and the national Greek Census. Particularly, the Fay-Herriot model was applied to produce estimates for two of the FGT measures (headcount ratio and poverty gap). Chapter 5 discusses the measurement of unemployment and focuses on the application of SAE methods for estimating unemployment in small geographical areas. Chapter 6 presents, analyzes, and interprets the results of the application of SAE methods for the estimation of unemployment in Greece at NUTS 3 level at two different times, in 2009 and 2013, combining data from the EU-SILC survey and the national Greek Census. Specifically, the Fay-Herriot model was applied to produce estimates for the unemployment rate. Finally, Chapter 7 contains some concluding remarks and suggestions for further work that can be done.

2. A review of Small Area Methods

2.1 Introduction

Small area estimation (SAE) procedures date back to the 1980s. Purcell and Kish (1980) described the estimation procedure for small areas using Census and administrative data. An extra step was taken by Ghosh and Rao (1994). The work of Sarndall (1984) and the monograph of Rao (2003) are considered crucial for the development of small areas statistics. Also, descriptions of the SAE theory can be found in the reviews of Rao (1999), Pfeffermann (2002; 2013), Jiang and Lahiri (2006), Datta (2009), Lehtonen and Veiganen (2009) and Guadarrama, Molina and Rao (2014). Rao (2003) in his monograph gives the following basic definitions for SAE methods.

> *Area or domain.* Examples of areas include a geographical region (e.g. a state, county, municipality, metropolitan area, health service area, etc.) or a socio-demographic group (e.g., a specific age-sex-race group within a large geographic area) or other subpopulations (e.g., the set of business firms belonging to a Census division by industry group) (Rao, 2003, p.1).
> *Direct Domain Estimator***.** A domain estimator is referred as direct if it is based only on the domain-specific sample data, that is it uses values of the variable of interest, y, only from the sample units in the domain.
> *Small Area***.** An area is regarded as small if the domain-specific sample is not large enough to support direct estimates of adequate precision (Rao, 2003, p.1).
> *Indirect Estimators.* Indirect estimators borrow strength by using values of the variable of interest, y, from related areas and/or time periods and thus increase the overall effective sample size and precision. These values are brought into the estimation process through a linking model (either implicit or explicit) based on auxiliary data such as recent Census and administrative records (Rao, 2003, p.2).

According to Pfeffermann (2002), the problem of SAE is twofold. The first problem is how to produce reliable estimates of characteristics of interest such as means, counts, quantiles, etc., for small areas or domains for which only small samples or no samples are available, and the second problem is how to assess the estimation error. Having only a small sample in each area, the only possible solution to the estimation problem is to borrow information from other related data sets, either from other 'similar' areas or from previous occasions. The different approaches to borrowing strength from information other than the observed values of the target variable in each small domain are (Elazar, 2004),

- cross-sectional (from other areas),

- using auxiliary data
- exploiting spatial relationship
- using over time relationship

Falorsi and Solari (2014) noted that cross-sectional is a way to borrow strength using values assumed by the target variable in all the areas included in the broad domain. This assumes that all the areas have a common mean value of the target variable. A second way to borrow strength is to divide the population into sub-groups according to one or more auxiliary items of information (such as age, income, etc.) and to assume a linear relationship between the target variable and the set of covariates. In this case we assume common mean values for all the domains within each sub-group. Another way to borrow strength is to use spatial information in the estimation process considering that units from the closest geographical areas are more important in terms of the auxiliary information they give. The last way to borrow strength from other sources of data is using over time relationship. In this case, information from previous surveys occasions or time is used. The first three approaches increase the effective domain sample size exploiting all the sampling information coming from the units belonging to the broad domain while the fourth approach increase the domain sample size using the sampling information coming from the units observed from previous survey occasions and within the target domain. The four approaches described above can be combined, defining in this way a classification of SAE methods as follows (Falorsi and Solari, 2014):

- methods involving spatial smoothing, using data of all the small domains for only one survey time
- methods involving temporal smoothing, using data for only the small domain of interest for several survey occasions
- methods involving spatial and temporal smoothing, using data collected for all the domains at different survey times

As noted by Pfeffermann (2002) the three classes of methods can be further divided according to the inferential approach. Based on this approach, traditionally there are two types of SAE, **direct** and **indirect** estimation. Direct SAE is based on survey design, so target parameters are considered unknown but fixed quantities. Bias, variance and other properties of estimators are evaluated under the randomization (design based) distribution. The randomization distribution of an estimator is the distribution over all samples that may be selected from the target population of interest. It is based on the sampling design used for sample selection and the population measurements being considered constant values. On the other hand, the indirect SAE uses model-based methods. Model-based methods depend on the selected sample, the target parameters are random variables, and the inference is based on the underlying model. This model (either implicit or explicit) provides a link to related areas using supplementary information related to the variables (Rao, 2003).

Furthermore, direct SAE is classified into three estimators called the H-T estimator, generalized regression (GREG) estimator and modified direct estimator. In addition, indirect SAE can be classified according to the linking model, if this is implicit or explicit (Rao, 2003). SAE based on implicit models include synthetic estimators and composite estimators. On the other hand, SAE based on explicit models are classified into aggregate level (area level) models and unit level models. Also, there are three standard basic methodologies for estimating explicit models based on SAE (Harding and Rahman (2017)). These are Empirical Best Linear Unbiased Prediction (EBLUP), Empirical Bayes (EB) and Hierarchical Bayes (HB).

A summary diagram of overall methodologies for SAE is presented in Figure 2.1.1.

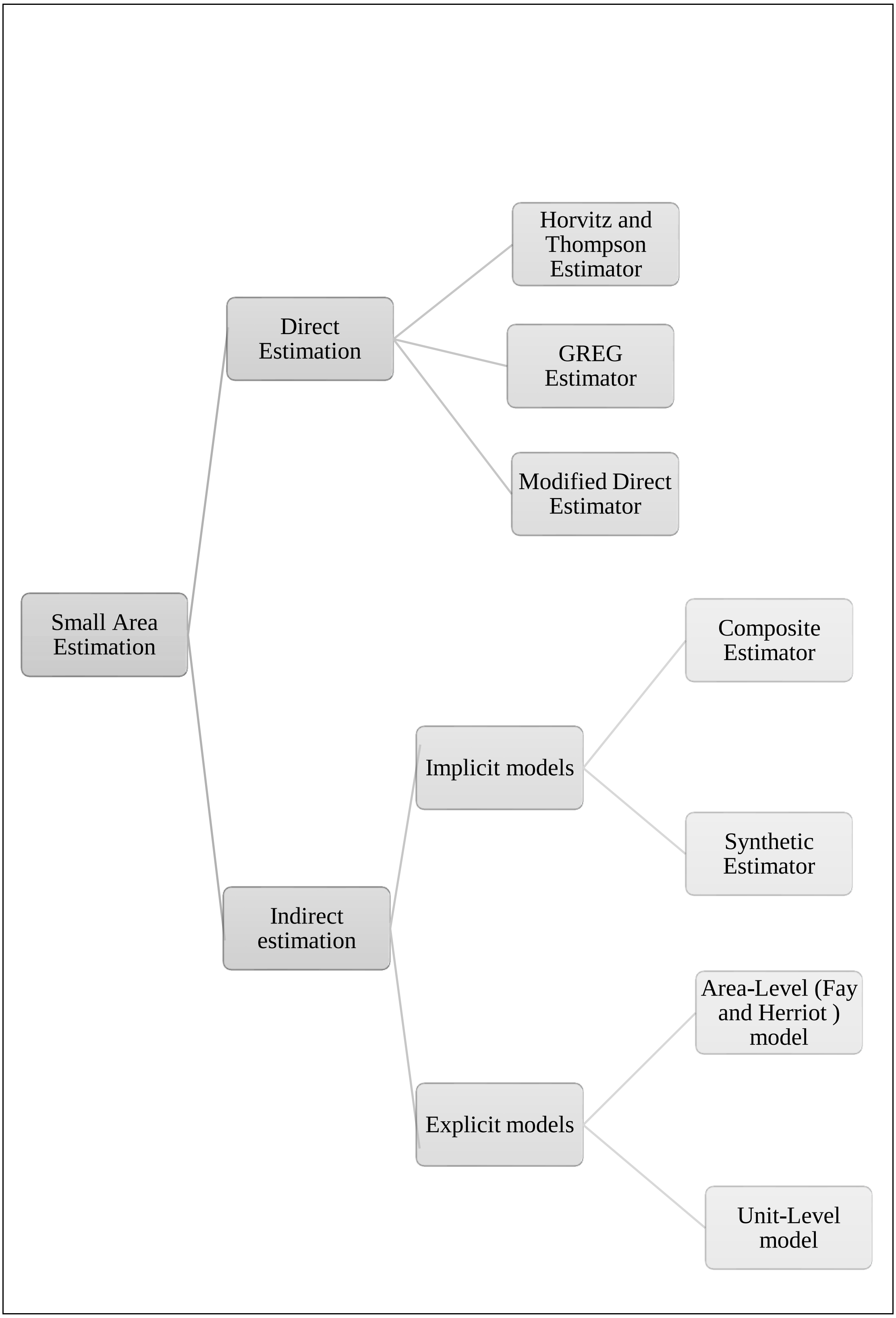

Figure 2.1.1 Classification of SAE methods

2.2 Direct estimators

A *direct* estimator is one that uses values of the variable of interest, y, only from the sample units in the domain of interest. Also, a direct estimator may use known auxiliary information, such as the total of an auxiliary variable, x, related to the variable of interest, y. Usually a direct estimator is design based. Design based estimators make use of survey weights and all the relevant inferences depend on the probability distribution resulting from the sample design, with the population measurements considered as fixed values (Rao, 2003). Common direct estimators applied to SAE include the Horvitz–Thompson (H-T) estimator (Cochran, 1977), GREG estimator (Sardnal, Swensson, and Wretman, 1992) and modified direct estimator (Rao, 2003; Rao and Molina, 2015).

2.2.1 Horvitz–Thompson (H-T) Estimator. Consider a finite population $U = \{1,2,..,j,...,N\}$ and a sample $s \subseteq U$ is drawn from U with a given probability sampling design, $\mathrm{p}(.)$. That is, $\mathrm{p}(s)$ is the probability that *s* is selected. Let y_j be the value of the variable of interest, y, for the j-th population element. The objective is to estimate the population total Y$= \sum_U y_j$, where $\sum_U$ denotes summation over the population elements j. Design weights $w_j(s)$ play an important role in constructing design-based estimators $\hat{Y}$ of Y. These basic weights may depend both on s and the element j $(j \in s)$. An important choice is $w_j(s) = \frac{1}{\pi_j}$, where $\pi_j = \sum_{\{s:j\in s\}} p(s)$, $j = 1,2,...,N$ are the inclusion probabilities[5] and $\{s: j \in s\}$ denotes summation over all samples *s* containing the element j. The weight w_j may be interpreted as the number of elements in the population represented by the sample element j. The Horvitz-Thompson (H-T) estimator (Cochran, 1977) of the population total Y is:

$$\hat{Y}^{HT} = \sum_s w_j\, y_j \tag{2.2.1}$$

where $\sum_s$ denotes summation over $j \in s$.

If $\pi_j > 0$, $j = 1,2,...,N$ the H-T estimator of Y is a design-unbiased estimator (Cochran 1977, Theorem 9 A.5), that is, $E\left(\hat{Y}^{HT}\right) =$ Y$= \sum_U y_j$.

Consider now that population U contains U_i $(U_i \subset U)$ subpopulation for a small area (domain) $i = 1,...,P$ with N_i elements (may or may not be known) and

[5] π_j=probability that the *j*-th unit is in the sample

$\sum_{i=1}^{P} N_i = N$. The objective is to estimate the domain total $Y_i = \sum_{U_i} y_j$ or the domain mean $\overline{Y}_i = \frac{Y_i}{N_i}$. If y_i is binary (1 or 0), then $\overline{Y}_i$ reduces to the domain proportion P_i (for example the proportion in poverty in the i-th domain).

The H-T estimator of the population total and of the domain mean for the i-th small area, can respectively be defined as (Rao, 2003):

$$\hat{Y}_i^{HT} = \hat{Y}(y_i) = \sum_{j \in s_i} w_{ij} y_{ij} \tag{2.2.2}$$

and

$$\hat{\overline{Y}}_i^{HT} = \hat{\overline{Y}}(y_i) = \frac{1}{\hat{N}_i} \sum_{j \in s_i} w_{ij} y_{ij} \tag{2.2.3}$$

where $s_i \subset s$ denotes the sample of elements belonging to domain U_i and $\hat{N}_i = \sum_{j \in s_i} w_{ij}$. Note that indicator i refers to small areas and j to individuals in the small area. Also, π_{ij} are the inclusion probability for individual j in area i and $w_{ij} = \frac{1}{\pi_{ij}}$ is the corresponding weight.

The estimator $\hat{Y}_i^{HT}$ is design-unbiased for Y_i if $\hat{Y}$ is design-unbiased for Y and it is also design-consistent[6] if the expected domain sample size is large. The H-T estimator is easy to calculate but when the sample size is inadequate (something common in the context of SAE problems) can be biased and unreliable.

2.2.2 Generalized Regression Estimator (GREG). Suppose now that y_j is the value of the variable of interest, y, for the j-th unit of the sample, with which is associated an observed auxiliary vector $\boldsymbol{x_j} = \left(x_{j1,\ldots\ldots,} x_{jp}\right)^T$ (p are the covariates). For the element $j \in s$, we observe $\left(\boldsymbol{x_j}, y_j\right)$. The population totals $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_p\right)^T$ of the auxiliary variable x are assumed to be known. This knowledge may come from one or more sources, such as Census or administrative data. The objective is to estimate the population total Y= $\sum_U y_j$. An estimator that makes efficient use of the auxiliary information is the generalized regression (GREG) estimator (or calibration estimator), proposed by Deville and Särndal (1992) and can be written as:

$$\hat{Y}^{GREG} = \hat{Y} + \left(\boldsymbol{X} - \hat{\boldsymbol{X}}\right)^T \hat{\boldsymbol{B}} \tag{2.2.4}$$

[6] An estimator $\hat{Y}$ of Y is said to be design-consistent if $\hat{Y}$ is design-unbiased and the design variance $V\left(\hat{Y}\right) = E\left[\hat{Y} - E\left(\hat{Y}\right)\right]^2$ of $\hat{Y}$ tends to zero as the sample size increase.

where $\hat{\boldsymbol{X}} = \sum_{j\in s} w_j x_j$ and $\hat{\boldsymbol{B}} = (B_1, B_2, \ldots, B_P)^T$ is the solution of the sample weighted least squares equations:

$$\left(\sum_s w_j q_j \boldsymbol{x_j}\, \boldsymbol{x_j}^T\right)\hat{\boldsymbol{B}} = \sum_s w_j q_j \boldsymbol{x_j}\, y_j \tag{2.2.5}$$

where $\hat{Y}$ and w_j defined in equation 2.2.1. Also, $1/q_j$ are known positive constants unrelated to w_j. In most applications this constant equals one (Sarndal et al., 1992). Consider now that domain specific auxiliary information is given in the form of p known domain totals $\boldsymbol{X_i} = \left(X_{i1}, \ldots, X_{ip}\right)^T$ for the i-th area and $\boldsymbol{x_{ij}} = \left(x_{ij1,\ldots\ldots}, x_{ijp}\right)^T$ is a vector with auxiliary information associated with y_{ij}. Also, y_{ij} is the value of the variable of interest y for the j-th population unit in the i-th area. For the element $j \in s_i$ we observe $\left(\boldsymbol{x_{ij}}, y_{ij}\right)$. Then the GREG estimator of the population total for the i-th small area can be defined as (Rao, 2003):

$$\hat{Y}_i^{GREG} = \hat{Y}_i + \left(\boldsymbol{X}_i - \hat{\boldsymbol{X}}_i\right)^T \hat{\boldsymbol{B}}_i \tag{2.2.6}$$

where $\hat{\boldsymbol{X}}_{\boldsymbol{i}} = \sum_{j\in s_i} w_j x_{ij}$ and $\hat{\boldsymbol{B}}_i = \left(\sum_{k\in s_i} w_{ik} q_{ik} \boldsymbol{x_{ik}} \boldsymbol{x_{ik}^T}\right)^T \left(\sum_{k\in s_i} w_{ik} q_{ik} \boldsymbol{x_{ik}} y_{ik}\right)$ is the sample weighted least square estimates of GREG.

The GREG estimator is a sum of the H-T estimator and a weighted difference between known totals and their H-T estimator. The GREG estimator is calibrated to the known auxiliary totals $\boldsymbol{X}$, that is $\hat{Y}^{GREG}(x) = \sum_s w_j^* x_j = \boldsymbol{X}$ (ESSnet, 2014b) and $w_j^* = w_j^*(s) = w_j(s) g_j(s)$ is the revised weight and is the product of the design weight $w_j(s)$ and the estimation weight $g_j(s) = 1 + \left(\mathbf{X} - \hat{\boldsymbol{X}}\right)^{\mathrm{T}} \left(\sum_s w_j q_j \mathbf{x}_{\mathrm{j}} \mathbf{x}_{\mathrm{j}}^T\right)^T q_j \mathbf{x}_{\mathbf{j}}$. In fact, GREG is a particular case of a calibration estimator when using the Euclidean distance. Moreover, all the calibration estimators can be asymptotically approximated by the GREG. A fundamental property of the GREG estimator is that it is nearly design unbiased[7] but not consistent because of high residuals (Särndal et al., 1992). Moreover, the evaluation of a GREG estimator variance is based on the variance of the residuals $(y_j - \hat{y}_j)$ (Särndal et al., 1992). The result of this, is that the better the fit of the linear working model the lower the variance of the GREG estimator and therefore the higher its accuracy. On the contrary, if the model underlying the GREG estimator is not appropriate for the target variable, a too large variation of weights may increase the

[7] Even being asymptotically unbiased, bias can be introduced if the sample size is too small.

variance with respect to the H-T estimator. Other disadvantages of the GREG estimator are that it produces negative weights (Rao, 2003) and can be very sensitive to the presence of outliers.

2.2.3 Modified Direct Estimator. Consider now that we replace $\hat{\boldsymbol{B}}_i$ in (2.2.6) by the overall regression coefficient $\hat{\boldsymbol{B}}$, given by (2.2.5). Then the modified direct estimator of the population total for the *i*-th small area is given by (Rao, 2003):

$$\hat{Y}_i^{MDE} = \hat{Y}_i + \left(\boldsymbol{X}_i - \hat{\boldsymbol{X}}_i\right)^T \hat{\boldsymbol{B}} \tag{2.2.7}$$

The modified direct estimator uses *y*-values from outside the domain but remains design-unbiased or approximately design-unbiased as the overall sample size increases, even if the domain sample size is small. This estimator is also known as the modified GREG estimator or the survey regression estimator, and even though it borrows strength for estimating the regression coefficient, it does not increase the effective sample size, something that indirect estimators do (Rao, 2003).

2.2.4 A Comparison of Direct Estimators. Direct estimators can perform well if the sample size is large enough (Harding and Rahman, 2017). Also, auxiliary information can be used to reduce the variance of the estimates. Although the theory and formulas of the direct small area estimators are very simple and straightforward, in the context of small area estimation direct estimators lead to unacceptably large standard errors because of very small samples from the small area of interest or sometimes of the complete lack of sample. In this case, direct estimators are statistically unreliable which makes it necessary to find indirect estimators that increase the effective sample size and thus decrease the standard error (Rao, 2003). Table 2.2.1 summarizes advantages and disadvantages of H-T, GREG and Modified Direct estimators and presents the recommended use of these estimators.

Table 2.2.1 A comparison of Direct Estimators

Estimator	Advantages	Disadvantages	Recommended use
H-T estimator	• Easy to calculate and it is design-unbiased for large samples	• Under inadequate sample size it can be biased and unreliable.	• Only if the sample size is large enough
GREG Estimator	• Can be used to reduce the variance of the estimates if a strong correlation between the target variable and the auxiliary variables exists. • Allows to calibrate to the known population totals of the auxiliary variables x. • It is approximately design unbiased	• Not consistent because of high residuals • Can be negative in some cases • Can introduce a large variation in weights that can cause an increase in variance • Even though asymptotically unbiased, bias can be introduced if sample size is too small • Can be very sensitive to presence of outliers	• When there is a linear relationship between target y and covariate variables x • Auxiliary information is available at unit or domain level • Sample size is large enough
Modified direct estimator	• It is design unbiased and uses overall aggregated data for coefficient estimation	• Borrows strength from the overall data but cannot increase the effective sample size	• When the overall sample size is large and reliable • Auxiliary information is available at unit or domain level

2.3 Indirect estimators

An indirect estimator borrow strength by using values of the variable of interest, y, from related areas and/or time periods and thus increases the overall effective sample size and precision. These values are brought into the estimation process through a linking model (either implicit or explicit) based on auxiliary data such as recent Census and administrative records (Rao, 2003). According to Schaible (1996), three types of indirect estimators can be identified: “area indirect”, “time indirect” and “area and time indirect”. An area indirect estimator makes use of y-values from another domain but not from another time period. A time indirect estimator uses y-values from another time period for the area of interest but not from another area. On the other hand, an area and time indirect estimator uses y-values from another area as well as another time period. Also, indirect SAE can be classified according to the linking model, if it is implicit or explicit (Rao, 2003). On the one hand implicit models (also called fixed effect models) assume that “the interdomain variability in the response variable can be explained entirely in terms of corresponding variability in the auxiliary information" (Chambers, 2003, p. 2) and on the other explicit models (also called area-specific random effects models) assume that "unexplained domain specific variability remains even after accounting for the auxiliary information" (Chambers, 2003, p. 2). Estimators based on implicit models are the Synthetic and Composite estimators while estimators based on explicit models are the Area-Level and Unit-Level Model estimators. An important issue in indirect SAE is the availability of good auxiliary data and determination of suitable linking models in order to increase the effective sample size and thus decrease the standard error.

2.3.1 Implicit Models. Implicit small area models provide a link to related small areas through supplementary data from Census and/or administrative records. These models (also called fixed effects models) "explain inter-domain variation in the response variable of interest, entirely in terms of variation in known factors" (Chambers, 2003, p. 2). Generally speaking, this approach includes two statistical techniques of indirect estimation — which are synthetic and composite.

2.3.1.1 Synthetic Estimator. According to Gonzalez (1973) "an estimator should be synthetic when a reliable direct estimator for a large area is used to derive an indirect estimator for a small area belonging to the large area under the assumption that all small areas have the same characteristics as the large area". Synthetic estimator uses both survey and auxiliary data from outside as well as within the domain of interest. A properly chosen model brings this auxiliary information in to the estimation process. Auxiliary information can be from different sources, such as sample survey data, Census data or administrative records.

Suppose that the only available information is the population size N_i in the i-th small area ($i = 1,2, \dots, P$). Then a synthetic estimator of the population total $\hat{Y}_i$ in the i-th small area is given by,

$$\hat{Y}_i^{synth} = N_i \frac{\sum_{j\in s} w_j y_j}{\sum_{j\in s} w_j} \tag{2.3.1}$$

where s, w_j, y_j is the same notation as in previous sections. The above estimator is also called Broad Area Ratio Estimator (BARE), (ESSnet, 2012c).

If domain-specific auxiliary information is available in the form of known totals X_i, then the regression-synthetic estimator of the population total $\hat{Y}_i$in the i-th small area (Rao, 2003) is given by:

$$\hat{Y}_i^{reg-synth} = X_i^T \hat{B} \tag{2.3.2}$$

where $\hat{B}$ is given by (2.2.5).

Note that in both cases the sample data are not available for the area of interest. The only information required is the local covariate totals or means and the value of $\hat{B}$, which is based on data from a wide area.

A special case of (2.3.2) is the ratio-synthetic estimator (Rao, 2003) in the case of a single auxiliary variable *x* and it is given by:

$$\hat{Y}_i^{ratio-synth} = X_i \frac{\hat{Y}}{\hat{X}} \tag{2.3.3}$$

where X_i is the known total value for the variable x for the *i*-th small area and $\hat{X} = \sum_{j \in s} w_j x_j$ is the direct survey estimate of the total of the only auxiliary variable at the broad area.

As synthetic estimator is easy to calculate and is unbiased if the assumption that small areas have the same characteristics as the broad area, is satisfied (assumption of homogeneity). On the other hand, it can be heavily biased if some small areas within the broad area have specific characteristics. Also, the choice of good auxiliary information, is very important. If the auxiliary information is not very predictive for the target variable, then predicted area means are pulled too much towards the general sample average (ESSnet, 2014).

We notice that the above formulation of synthetic estimator is based on obtaining the most reliable direct H-T estimator for the wide area and using it to export an estimator for the small area. There is an exchange between bias and precision in the direct and synthetic estimators. Direct estimators have little or no bias, but possibly low precision while the synthetic estimator is biased but more precise (ESSnet, 2012c). Finally, it is worth noting that synthetic estimators can be applied to general sampling designs or surveys where a sample design is not present.

2.3.1.2 Composite Estimation. In many cases, the direct estimator is not taken into account because of its large variance and the synthetic estimator give unacceptable results because of bias. A solution to this problem can be given by an alternative estimator which is as a weighted sum of direct and synthetic estimators. This type of estimator is commonly known as a composite estimator. According to Ghosh and Rao

(1994), a composite estimator is a natural way to balance the potential bias of a synthetic estimator against the instability of a direct estimator by choosing an appropriate weight.

Let, $\hat{Y}_{i1}$ and $\hat{Y}_{i2}$ be the direct and synthetic estimator respectively, of the small area total Y_i. Then, the composite estimator of population total Y_i for a small area i can be defined as:

$$\hat{Y}_i^{comp} = \varphi_i \hat{Y}_{i1} + (1 - \varphi_i)\hat{Y}_{i2} \tag{2.3.4}$$

for a suitably chosen weight φ_i $(0 \leq \varphi_i \leq 1)$. Many of the estimators, both design and model based, have this basic form. Note that the weight of the direct estimator $\hat{Y}_{i1}$ grows as the sample size increases whereas the weight of the synthetic estimator $\hat{Y}_{i2}$ grows as the sample size gets smaller. A suitable choice of the weight φ_i is crucial, can be done in different ways and ranges from simple to optimal weight. One of the most common solutions is to take $\varphi_i = \frac{n_i}{N_i}$, where n_i is the sample size for domain i and N_i is the population size for domain i. Alternatively, the optimal weight can be obtained by minimizing the design mean square error (MSE) of the composite estimator, $\hat{Y}_i^{comp}$, with respect to φ_i, under the assumption that the covariance factor of $\hat{Y}_{i1}$ and $\hat{Y}_{i2}$ is small in comparison to the MSE of $\hat{Y}_{i2}$ (Rao, 2003). Now if $\varphi_i{}^*$ represents the optimal weight, then it can be defined as:

$$\varphi_i{}^* = \frac{MSE\left(\hat{Y}_{i2}\right)}{\left[MSE\left(\hat{Y}_{i1}\right) + MSE\left(\hat{Y}_{i2}\right)\right]} \tag{2.3.5}$$

Rao (2003, section 4.3) provides different types of composite estimator under a design-based approach, including the Sample Size Dependent Estimator (SSD) and the James-Stein Method as well as examples of their applications. Some other ways of finding φ_i are discussed in Ghosh and Rao (1994), Holmoy and Thomsen (1998) and Singh, Gambino and Mantel (1993).

Composite estimators can be useful in surveys in which the sample sizes vary considerably among the domains of interest. When the sample size is quite large the direct estimator is valuable. On the other hand, when the sample size is small or even equal to zero synthetic estimators are more valuable. A composite estimator balances these two cases in order to avoid switching from the one estimator to the other, so is less biased than synthetic estimate (ESSnet, 2014c).

2.3.1.3 A Comparison of Indirect Estimators based on Implicit models. As discussed in the previous sections, when the sample size for domains of interest is too small or even equal to zero, direct estimators based only on specific domain sample data are insufficient because of high variability and low precision. The solution to this problem can be given by Small Area Estimation methods. One class of these methods are Indirect Estimators (Synthetic and Composite) based on Implicit Models. Synthetic estimators are easy to calculate and can be applied even in domains with no sample. Also, if the assumption of homogeneity is satisfied then the synthetic estimator is unbiased. Composite estimators are also easy to implement and give a compromise between the large variance of a direct estimator and the bias of a synthetic estimator. In addition, it is worthy of note that although these indirect estimators have better precision and reliability compared to the direct small area estimators, each of these has also disadvantages. A comparative summary for Synthetic and Composite estimators is presented in Table 2.3.1.

Table 2.3.1 A comparison of Indirect Estimators (Implicit models)

Indirect Estimators			
Implicit Models	**Advantages**	**Disadvantages**	**Recommended use**
Synthetic estimator	• Easy to calculate. • It can be applied to a domain of interest even when there is no unit from that domain in the sample. • If all the small areas have the same characteristics as the broad area, then the synthetic estimator is unbiased.	• If the small areas do not have the same characteristics as the broad area, then the estimator may be strongly biased. • If the auxiliary information used for synthetic estimators is not very predictive for the target variable, then predicted area means are pulled too much towards the general sample average.	• When a domain in the sample is not represented at all or there are only a few sampled units in specific domains. • Can be used even when sampling was not involved.
Composite estimator	• Is easy to implement and understand by users. • Balances the potential bias of the synthetic estimator against the instability of the direct estimator	• It is not always easy to establish the value of the weight φ_i. • Even though its bias is less than that of synthetic estimator it still present.	• Can be useful in surveys in which the sample sizes vary considerably among the domains of interest.

2.3.2 Explicit Models. Explicit linking models in the literature are called "small area models". These models are mixed models that incorporate domain specific random effects. As noted by Saei and Chambers (2003, p.2) "these models assume that "unexplained" domain specific variability remains even after accounting for the auxiliary information". Available small area models can be classified into two broad types:

- Basic area level models in which the information about the response variable is available only at small area level.
- Basic unit level models in which information about the response variable is available at unit level.

2.3.2.1 Basic Area Level Model (Fay and Herriot model). The basic area level model assumes that the true mean values of the parameter of interest are linearly related to the values of some auxiliary variables measured at the area level through a linear regression model (Molina and Morales, 2009**.** The model is constructed in two stages. In the first stage a sampling model is used for the direct estimates and in the second stage a linking model is used for the parameters of interest. The sampling model includes the direct estimator of the survey and the corresponding sampling variance while the linking model relates the parameter of interest with a regression model with area-specific random effects (Kordos, 2016).

In more detail, suppose that $\boldsymbol{x_i} = \left(x_{1i}, x_{2i}, \dots, x_{pi}\right)$ is a vector of covariates (area-specific auxiliary data) for domain i and θ_i is the parameter to be estimated for this domain *i*, with $i = 1,2, \dots D$. A linear relationship between θ_i and covariates $\boldsymbol{x_i}$ (whose values are known for each domain) is assumed, that is:

$$\theta_i = \boldsymbol{x_i}^T \boldsymbol{\beta} + u_i,\ \ i = 1,2, \dots D \tag{2.3.6}$$

where $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \dots, \beta_p\right)$ is the $p \times 1$ vector of regression coefficients and u_i's are area specific random effects which are assumed to be independent and identically distributed with mean zero $(E_D(u_i) = 0)$ and variance $V_D(u_i) = \sigma_u{}^2$, $(u_i \overset{iid}{\sim} (o, \sigma_u{}^2))$. The random effects account for the extra variability not explained by the auxiliary variables in the model.

Consider now that a design unbiased direct estimator $\overset{\wedge}{\theta_i}$ and the correspondent variance are available for each area *i*. Then the following model is assumed:

$$\overset{\wedge}{\theta_i} = \theta_i + e_i\ , i = 1,2, \dots D \tag{2.3.7}$$

where e_i is the sampling error associated with the direct estimates of each small area i. Sampling errors e_i are supposed to be independent with $E(e_i|\theta_i) = 0$ and $V(e_i|\theta_i) = \psi_i$, where sampling variances ψ_i are assumed to be known ($e_i \overset{ind}{\sim} (0, \psi_i)$). Combining (2.3.6) and (2.3.7) we obtain the model (Rao and Molina, 2015):

$$\hat{\theta}_i = \boldsymbol{x_i}^T \boldsymbol{\beta} + u_i + e_i \ , \ i = 1,2,\dots D \tag{2.3.8}$$

Model (2.3.8) is a special case of a linear mixed model and is also called the Fay-Herriot model[8]. Note that (2.3.8) involves design-induced errors e_i as well as model errors u_i which are considered independent. Normality for both u_i and e_i is also often assumed for the MSE estimation, even if this assumption is not necessary for estimating the parameters.

Furthermore, the assumption of known sampling variances ψ_i may be restrictive and the assumption of $E(e_i|\theta_i) = 0$ may not be tenable if the small area sample size n_i is very small and θ_i is a nonlinear function of the small area total Y_i (Rao, 2003). However, if information at unit level is available, then under the hypothesis of homoscedasticity of the sampling errors, the variance ψ_i can be estimated from a unit level model or a generalized variance function (Wolter, 2007). But this affects the Mean Square Error (MSE) of the predicted domain values (Bell, 1999). Wang and Fuller (2003), You and Chapman (2006), Gonzalez-Manteiga, Lombardia, Molina, Morales and Santamaria (2010), considered the situation where the sampling variances ψ_i are unknown and modelled separately by direct estimators.

In practice the fixed effects parameters $\boldsymbol{\beta}$ and the variance, ${\sigma_u}^2$, of the random effects are generally unknown. In order to compute the estimates, under a model based small area approach, three methods can be used, known as the empirical best linear unbiased prediction (EBLUP), Empirical Bayes (EB) and hierarchical Bayes (HB) procedures. An overview of these procedures is presented in section 2.4.

The Fay-Herriot model (2.3.8) is used widely in practice to obtain reliable model-based estimates for small areas. In particular, it has been applied by the S.A.I.P.E project since 1993 to produce on an annual basis, model-based county estimates of poor school-age children in the United States[9]. Also, the Fay-Herriot model is used for the

[8] Fay and Herriot (1979) were the first to use a model like (2.3.8) for small area estimation. Specifically, they used such a model to estimate the Per Capita Income (PCI) for small places in the United States with population less than 1,000.
[9] More details are given in paragraph 3.5.

Census undercount by Statistics Canada (Dick, 1995) with $\theta_i = \frac{T_i}{C_i}$, where T_i is the true (unknown) count, C_i is the Census count and i denotes province, age and sex combination.

The basic area level model assumes symmetry of the distribution of area random effects, which may not hold in practice, for example, in business surveys. In this case, if transformation of variables does not suffice to reduce skewness, advanced methods have been proposed (Chandra and Chambers, 2007).

Moreover, there are many extensions of the basic area-level model that allow time and spatial correlation patterns to be taken into account, as well as correlated sampling errors, which must be taken into account when conducting repeated surveys with rotating samples. Mixed area-level models with different area-time effect correlation structures have been considered in Eurarea (2004). Furthermore, Tiller (1991) uses a time series approach to model the true unemployment rate. To achieve this, he defines an ARMA model to formulate the time correlation of sampling errors of the estimates of this rate. In addition, Rao and Yu (1994) have used an AR model as a generalization of the area mixed model. The AR model can combine cross-sectional data with information observed in previous stage of the survey. Singh, Gambino and Mantel (1994) used a different approach to handle the correlation between parameters in relation to area and time. In this context, Pfeffermann and Tiller (2006) have proposed some restrictions so that the small area estimates sum to the direct estimates at an aggregated level. Moreover, Pratesi and Salvati (2008) proposed, within the framework of spatial autocorrelation between areas, a spatial EBLUP estimator based on an area linear mixed model. In this model the random area effect of each area is correlated with the random effects of its neighbors. You and Rao (2002), taking into account that sometimes the target variable is not a linear function of population total or mean, proposed an extension of the basic mixed areal level model to deal with the unmatched sampling and linking models. Furthermore, multivariate mixed area level models have been developed for a multivariate response (Datta, Ghosh, Nangia and Natarajan, 1996; Rao, 2003). This type of model reduces the MSE of the small area estimates by considering the correlations with the other variables.

2.3.2.2 Basic Unit Level Model. The unit level model relates the unit values of the study variable to unit-specific auxiliary variables. This model is a linear combination of the direct information and a regression synthetic prediction of nonsampled units (Pfeffermann, 2002). The fixed part of model links, for each unit, the target values to some known auxiliary variables. The area specific random effects are introduced in order to take into account the correlation among the units with each small area (between area variation). The basic unit level linear mixed model is the nested error regression model formulated by Battese, Harter and Fuller (1988) and can be expressed as follows:

$$y_{ij} = \boldsymbol{x_{ij}}^T\boldsymbol{\beta} + u_i + e_{ij}, j = 1,2,\dots,N_i \text{ and } i = 1,2,\dots\dots,D \qquad (2.3.9)$$

where y_{ij} is the variable of interest and $\boldsymbol{x_{ij}} = \left(x_{ij1}, x_{ij2},\dots,x_{ijp}\right)^T$ represents the unit-specific auxiliary data for population element j in small area i. This auxiliary data is available for areas $i = 1,2,\dots\dots,D$ and $j = 1,2,\dots,N_i$, as N_i is the number of population units in the i-th area.

The area-specific effects u_i are assumed to be normal, independent, and identically distributed random variables with mean zero and variance ${\sigma_u}^2$, ($u_i \overset{iid}{\sim} N(o,{\sigma_u}^2)$).

The sampling errors e_{ij} are independent of u_i's and follow independent and identical normal distribution with mean zero and variance σ_e^2, ($e_{ij} \overset{iid}{\sim} N(o,{\sigma_e}^2)$). Assuming that a non-informative sampling design, like simple random sampling, has been used at the sampling stage, the same model assumed for the population values can be applied for the sample units (Rao, 2003). For this, the model for the sample units can be written, using a matrix formulation, as:

$$\boldsymbol{y_i} = \boldsymbol{X_i\beta} + \boldsymbol{u_i 1_i} + \boldsymbol{e_i}, i = 1,2,\dots\dots,D \qquad (2.3.10)$$

where $\boldsymbol{X_i}$ is $N_i \times p$ vector and $\boldsymbol{y_i}$, $\boldsymbol{1_i}$, $\boldsymbol{e_i}$ are $N_i \times 1$ vectors and $\boldsymbol{1_i} = (1,\dots,1)^T$.

To obtain the small area estimates based on the above model one of the following three approaches can be used (Pfeffermann, 2002): a predictive, an empirical or a hierarchical Bayesian approach. Under the predictive approach, the Best Linear Unbiased Predictor (BLUP) is obtained by minimizing the quadratic loss in the linear unbiased estimator class (Henderson, 1975). Since the BLUP estimator depends on variance components ${\sigma_v}^2$and σ_e^2, that are usually unknown, their estimates need to be computed. This can be achieved in different ways, for example by means of Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) methods (Cressie, 1992).

The nested error unit level regression model (2.3.9) was first used to estimate county crop areas in the United States (Battese et al. 1988) combining sample survey data with satellite information. In particular, the purpose was to estimate the area under corn and soybeans for each of the D=12 counties in North-Central Iowa. A sample was taken from the small areas into which each county was divided. The areas, in the sample, under corn and soybeans were confirmed by interviewing farm operators. Auxiliary data, in the form of numbers of pixels classified as corn and soybeans, were obtained for all the small areas in each county using satellite information. This auxiliary data was in the form $x_{ij} = \left(1, x_{ij1}, x_{ij2}\right)^T$ where x_{ij1}= number of pixels classified as corn and x_{ij2}=number of pixels classified as soybeans in the *j*-th area of the *i*- th county. The model proposed by Battese et al. (1988) was:

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + u_i + e_{ij} \qquad (2.3.11)$$

where y_{ij}=number of hectares of corn (or soybeans) in the *j*-th area of the *i*-th county.

As already mentioned above, the basic unit level model assumes sampling designs that use the auxiliary information x_{ij} in the selection of the samples s_i. But when such an informative design is used, the inclusion probabilities of sampling units depend on the values of the target variable and then the model which holds for the sample data is different from the model assumed for the population data. This can produce severe bias in the predictor of the variable of interest. In order to overcome this problem several extensions of the basic unit level model (2.3.9) have been proposed. For instance, Stukel and Rao (1999) proposed a two-fold nested error regression model for data collected from a stratified two-stage sampling. Pfeffermann and Sverchkov (2007) proposed under a unit level mixed model an estimator that takes into account the sampling weights distribution within the sampled areas. Also, Prasad and Rao (1999), proposed a Pseudo EBLUP estimator starting from the basic unit linear mixed model. This model is weighted through normalized weights, achieving in this way a survey-weighted aggregated area level model.

In addition, the basic unit level model assumes symmetry (or normality) of the distribution of random effects, which may not hold in practice, like in business surveys. In this case, if transformation of variables does not suffice to reduce skewness, advanced methods have been proposed. For instance, Chambers and Tzavidis (2006) and Tzavidis, Marchetti and Chambers (2010) proposed M-quantile models. In the case of M-quantile models the assumption of symmetry (or normality) on the random effects can be relaxed.

Moreover, Datta, Day, and Basawa (1999), applied a multivariate nested error regression model that takes into account the correlation between the characteristics under study. This model makes it possible to estimate more than one small area parameter of interest.

Also, since linear unit level mixed models apply only for continuous observations, generalized linear mixed models (GLMM) (Jiang and Lahiri, 2006; Ghosh, Natarajan, Walter and Kim, 1999)) can be considered in the case of categorical dependent variables. These models allow us to deal with discrete responses that are quite common in practice. GLMM models have been mainly developed using a Bayesian approach to inference. For instance, MacGibbon and Tomberlin (1989) used a logistic regression model with random area-specific effects in the binary response data. Malec, Sedransk, Moriarity and LeClere (1997) considered a different logistic regression model with random regression coefficients, which was applied to estimate the proportion of persons in a state or substate who have visited a physician in the past year, using the data from the U.S. National Health Interview Survey.

2.3.2.3 A Comparison of Indirect Estimators based on explicit models. It is generally expected that when indirect estimators are to be used, they should be based on explicit small area models. Such models define the way that the related data are incorporated in the estimation procedure. Explicit small area models include a range of statistical models that are classified into two broad types, *area level* and *unit level* models. According to Rao and Molina (2015), the use of explicit models offers several advantages. Firstly, model diagnostics can be used to find a suitable model that fits the data well and area-specific measures of precision can be associated with each small area estimate. Also, linear mixed models can be considered as well as nonlinear models such as logistic regression and generalized linear models with random effects. Another advantage is that small area models can cope with complex data structures, such as spatial dependence and time series structures.

It is important to note, that availability of good auxiliary data and determination of suitable linking models are crucial to the formation of small area models. If the assumed model does not provide a good fit to the data, the small area estimator will be model biased which in term can lead to erroneous inferences (Whitworth, 2013).

In general, unit-level models are preferred over area-level models (Hidiroglou and You, 2016). The main reason is that the auxiliary information available at the unit level provides more explanatory power to the model. The decision between a unit level and an area level model is mainly determined by whether most of the variation occurs at the unit level or more broadly at the area level. When there is a substantially larger variability at unit level, a unit level model will more effectively partition the estimate of the variability between the levels, thus giving more accurate area precision estimates (ESSnet, 2012c). However, due to the difficulty in obtaining and matching auxiliary unit-level data to the survey respondents, the use of a unit-level model is not always possible. In this case, an area level model can be used.

An additional concern for the type of model preferred (area or unit level model) is the ecological fallacy. The ecological fallacy (Heady and Hennell, 2000) describes the fault of inferring individual behavior from aggregate data relationships. In the small area estimation context, the concern is that, regardless of whether the relationship is inferred at individual or at area level, it is applied at area level in order to determine the small area estimates. In this case the resulting dilemma is the following: Do we use area level auxiliary variables, even if unit level ones are available and can help explain the within area variability, or do we use unit level variables and risk incorrect area level estimates? One solution is to use both, the area and the unit level models. This will allow separate coefficients to be estimated for the area and unit levels of the same auxiliary variable. So, if an ecological effect is not present, the values of the coefficients in each pair of area and unit level variables will be similar.

A comparative summary for area and unit level models is presented in Table 2.3.2 according to their advantages, disadvantages, and recommended use.

Table 2.3.2 A comparison of Indirect Estimators (Explicit models)

Indirect Estimators			
Explicit Models	**Advantages**	**Disadvantages**	**Recommended use**
Basic Area Level Model (Fay and Herriot)	• The method increases the reliability of the estimates by introducing a linear relationship between the direct and known area level auxiliary variables. • Covariates are needed only at domain level. • The method is applied by U.S. Census for poverty estimation since 1993 and used by Statistics Canada for Census undercount estimation.	• If the model is not correctly specified, the estimator can be affected by bias. • Assumptions of normality with known variance might be untenable at small sample sizes.	• Can be applied for estimation when few or even no sample data[10] are available for one or more domains of interest. • Can be applied for estimation when a set of covariates with a strong relationship with the variable of interest is available. • The method can be applied when area level auxiliary data are available for sampled and not-sampled areas and the direct survey estimates are available for sampled areas.
Basic Unit Level Model	• The method is useful to improve the direct estimator if a set of covariates with a strong relationship with the target variable is available. • The extra information available at the unit level provides additional explanatory power. • Used successfully in many areas of agricultural statistics.	• If the model is not correctly specified, the estimator can be affected by severe bias. • The basic method does not consider the sampling strategy to select units. • The model assumes symmetry of the distribution of random effects, while may not hold in practice.	• The method can be applied when unit level auxiliary information is available. • The method can be applied when data are continuous and normally distributed otherwise, a transformation of the data may be required. • The method can be applied when a set of covariates with a strong relationship with the target variable is available.

[10] For domains with no data only synthetic estimates can be computed. Based on model (2.3.8), an estimator making use of only the regression component is given by the area level synthetic estimator: $\hat{\theta}_i^{synth-arealevel} = \boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}$. This estimator uses only the relationship between the target variable and the covariates and does not exploits the direct information.

2.4 Methods for Estimating Explicit Models

In order to produce the estimator and the corresponding measure of the accuracy of small area estimates, different inferential approaches can be applied. These can generally be classified into predictive and Bayesian methods of inference (Chambers, 2003).

The predictive approach is widely used in the context of SAE. A key assumption for its application is that the variances associated with random effects in the mixed models (variance components) are known. In order to estimate the target parameter, the Best Linear Unbiased Predictor (BLUP) is obtained (Henderson, 1975). In Best Linear Unbiased Predictor, "Best" refers to the minimum mean square error (MSE) of all linear unbiased predictors, "linear" refers to the linear combination of the predictor with the response variable values and "unbiased" means that the expected value of the prediction error is zero. In practice, the variance components are unknown and must be estimated from the data. The estimation for variance components can be obtained using Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) or Method of moments (Harville, 1997; Cressie, 1992). Using these estimated components in the BLUP estimator, the obtained estimator is called the Empirical Best Linear Unbiased Predictor (EBLUP) (Harville, 1991).

The Bayesian approach is based on the posterior probability function. The predictor of the target parameter is given by the mean of its posterior distribution. Under this approach Empirical Bayes (EB), Hierarchical Bayes (HB) estimation and inference methods can be applied. In the Empirical Bayes (EB) method the unknown model parameters are estimated from the marginal distribution of the data and then plugged into the Bayes predictor formula (Prasad and Rao, 1990). In the Hierarchical Bayes (HB) method, the fixed effect parameter $\boldsymbol{\beta}$ and the variance components are considered random, and a prior distribution of these parameters is assumed (Datta and Ghosh, 1991).

EBLUP, EB, and HB methods have played a significant role in studying various small area models for indirect SAE. A concise summary of these three statistical procedures is given below.

2.4.1 Empirical Best Linear Unbiased Predictor (EBLUP). To predict the random effects or mixed effects for a small area model, the best linear unbiased prediction (BLUP) approach is widely used. The BLUP method was originated by

Henderson (1950), and many authors have used it in different ways and in various fields. A comprehensive overview of the derivations of the BLUP estimator with useful examples and applications is provided in Robinson (1991), You and Rao (2002a) and Rao (2003).

Assume that the sample data follow the general linear mixed model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \tag{2.4.1}$$

where, $\boldsymbol{y}$ is the $n \times 1$ vector of sample observations, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known $n \times p$ and $n \times h$ matrices of full rank, and $\boldsymbol{v}$ and $\boldsymbol{e}$ are independently distributed with means zero and covariance matrices **G** and **R** depending on some variance parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)^T$. Also, $Var(\boldsymbol{y})$ denotes the variance-covariance matrix of $\boldsymbol{y}$ with $Var(\boldsymbol{y}) = \boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\delta}) = \boldsymbol{R} + \boldsymbol{ZGZ}^T$.

Consider a linear combination, $\mu = \boldsymbol{1}^T\boldsymbol{\beta} + \boldsymbol{m}^T\boldsymbol{v}$ of the regression parameters $\boldsymbol{\beta}$ and specified vectors, $\boldsymbol{1}$ and $\boldsymbol{m}$. A linear estimator of μ is of the form $\hat{\mu} = \boldsymbol{a}^T\boldsymbol{y} + \boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are known.

Suppose E denotes the expectation with respect to the model (2.4.1), the estimator $\hat{\mu}$ is model-unbiased for μ if $E(\hat{\mu}) = E(\mu)$. The MSE of $\hat{\mu}$ is given by:

$$MSE(\hat{\mu}) = E\left(\hat{\mu} - \mu\right)^2 \tag{2.4.2}$$

When $\hat{\mu}$ is unbiased for μ then the MSE reduces to the variance of the error $\hat{\mu} - \mu$, that is $MSE(\hat{\mu}) = Var\left(\hat{\mu} - \mu\right)$.

The objective is to find the BLUP estimator which minimizes the MSE in the class of linear unbiased estimators $\hat{\mu}$. For known $\boldsymbol{\delta}$, the BLUP estimator of μ is given by (Rao, 2003):

$$\tilde{\mu} = t(\boldsymbol{\delta}, \boldsymbol{y}) = \boldsymbol{1}^T\tilde{\boldsymbol{\beta}} + \boldsymbol{m}^T\tilde{\boldsymbol{v}} = \boldsymbol{1}^T\tilde{\boldsymbol{\beta}} + \boldsymbol{m}^T\boldsymbol{GZ}^T\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}\right) \tag{2.4.3}$$

where, $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{v}}(\boldsymbol{\delta}) = \boldsymbol{GZ}^T\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}\right)$.

The BLUP estimator $t(\boldsymbol{\delta}, \boldsymbol{y})$ given by (2.4.3) depends on the variance parameters $\boldsymbol{\delta}$ which in the usual BLUP approach are assumed to be known. However, in practice, $\boldsymbol{\delta}$ are usually unknown, and they are estimated from the sample data (You and Rao, 2002b). In these cases, a two-stage BLUP estimator can be developed, which is currently well known as the Empirical Best Linear Unbiased Prediction, EBLUP

(Ghosh and Rao, 1994; Pfeffermann, 2002; Rao, 2003). EBLUP estimator is obtained by replacing $\boldsymbol{\delta}$ by an estimator $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}(\boldsymbol{y})$, so it is $\hat{\mu}^{EBLUP} = t(\hat{\boldsymbol{\delta}}, \boldsymbol{y})$.

Kackar and Harville (1981) showed that a two-step approach in which variance components are first estimated and then used to estimate and predict fixed parameters and random components can lead to unbiased estimators. The two-stage estimator remains unbiased if the distribution of the data vector is symmetric about its expected value, the variance component estimators are translation-invariant and also are functions of the data vector. Kackar and Harville (1981) also showed that the maximum likelihood (ML) and residual maximum likelihood (REML) variance component estimators have the above properties.

Particularly for the basic area level model (using the notation from paragraph 2.3.2.1) the BLUP estimator for the parameter θ_i can be written as (Molina and Morales, 2009):

$$\hat{\theta}_i^{BLUP} = \gamma_i \hat{\theta}_i + (1-\gamma_i) x_i \tilde{\beta} \tag{2.4.4}$$

where $\hat{\theta}_i$ is the Fay-Herriot estimator given by (2.3.8), $\boldsymbol{x_i}$ is the vector of area-specific covariates for domain i and γ_i is the proportion of variance due to area specific random effects u_i that is given by the formula:

$$\gamma_i = \sigma_u^2/(\sigma_u^2 + \psi_i) \tag{2.4.5}$$

Also, $\tilde{\boldsymbol{\beta}}$ is the weighted least squares estimator of regression coefficients $\boldsymbol{\beta}$ given by

$$\tilde{\boldsymbol{\beta}} = (\textstyle\sum_{i=1}^{D} \gamma_i \boldsymbol{x_i}' \boldsymbol{x_i})^{-1} \textstyle\sum_{i=1}^{D} \gamma_i \boldsymbol{x_i}' \hat{\theta}_i \tag{2.4.6}$$

When the variance σ_u^2 is large in relation to the total variance then the BLUP in (2.4.4) approaches the direct estimator of the domain. Conversely when the variance σ_u^2 is small relative to the total variance then the BLUP is close to the synthetic regression estimator. In addition, as pointed out by Molina and Morales (2009), the BLUP estimator based on the Fay-Herriot model is at least as efficient as the direct estimator and more efficient for areas with larger sampling variances.

As already mentioned, σ_u^2 is usually unknown. ML or REML can be used as estimating methods of σ_u^2. REML gives a less biased estimator of σ_u^2 as accounts for the degrees of freedom due to the estimation of $\boldsymbol{\beta}$ (Rao and Molina, 2015). Replacing σ_u^2 by an estimator $\hat{\sigma}_u^2$ in (2.4.4) the Empirical BLUP (EBLUP) is obtained, given by (Rao, 2003):

$$\hat{\theta}_i^{EBLUP} = \hat{\gamma}_i\hat{\theta}_i + (1 - \hat{\gamma}_i)\boldsymbol{x_i}\hat{\boldsymbol{\beta}}, \tag{2.4.7}$$

where $\hat{\gamma}_i$ and $\hat{\boldsymbol{\beta}}$ are the values of γ_i and $\tilde{\boldsymbol{\beta}}$ when σ_u^2 is replaced by the estimator $\hat{\sigma}_u^2$.

2.4.2 Empirical Bayes and Hierarchical Bayes method. Another important method used to predict the random effects for a small area is the Bayesian approach. This approach includes the Empirical Bayes (EB) and Hierarchical Bayes (HB) methods that are widely used in small area estimation (Saei and Chambers, 2003).

Referring to the EB method (Morris, 1983) the data sample is used to estimate the joint prior and posterior distributions of the small area quantities of interest. These distributions usually depend on unknown model parameters. Common techniques such as ML or REML are used to estimate them. A major problem with the EB method is the need to take uncertainty into account when estimating the prior or posterior distributions. Among the various approaches that have been proposed to address this problem are the delta (Deely and Lindley, 1981) and bootstrap methods (Laird and Louis, 1987).

As far as the HB method is concerned, the modeling is done in stages and each stage is relatively simple and can be easily understood although the entire process of model fitting can be complicated (Saei and Chambers, 2003). The HB method assumes a joint prior distribution for the fixed effect parameter $\boldsymbol{\beta}$ and the variance components (these parameters are treated as random). Bayes Theorem is then used to determine the joint posterior distribution of the small area quantities of interest (Rao, 2003). For specified variance components the estimates of the small areas resulting from the HB method usually coincide with the BLUP and EB estimates. According to Ghosh and Rao (1994) HB estimators under some assumptions about the prior parameter distribution of the model have a lower MSE than the corresponding BLUP-based estimators. However, the computational complexity of the HB method is a clear disadvantage. Usually, simulation-based methods such as Markov Chain Monte Carlo (MCMC) (Brooks, 1988) are required to approximate the posterior distribution as well as the posterior means and variances of the small area quantities of interest (You and Rao, 2002b).

According to Rao (2003) the Bayesian approach has several advantages. One of them is that it easily includes different types of target variables such as binary or count data as well as more complex random effects structures, such as time or spatial correlation. Also, the uncertainty about model parameters is directly taken into account in the posterior distribution of the small area estimates. Furthermore, Bayesian methods can also deal with even high-complexity models as opposed to EBLUP estimators for which inference becomes more difficult as the complexity of the model increases. On the other hand, as mentioned above, the theory of Bayesian methods is usually considered to be more difficult than that of EBLUP. Finally, it is worth noting that many times EBLUP, EB and HB methods can provide the same result under specific assumptions (Pfeffermann, 2002).

2.5 Selecting auxiliary data

The selection of auxiliary data related to the variables of interest is crucial to the formation of indirect estimates. The success of any model-based method depends on how good the auxiliary data is, "good" in the sense that they are good predictors of the study variables (Rao, 2003).

The first problem to be concerned with selection of auxiliary variables is the availability of data. Usually, administrative data and Censuses are the best sources as their area-based values are free of sampling errors (although they could suffer from non-sampling errors). However, administrative and Census data are rarely available at individual level for confidentiality reasons, so it is necessary to use auxiliary data from other surveys. As noted by Scaible (1996), expanded access to auxiliary information through coordination and cooperation among different agencies is needed.

The second problem is which of the available auxiliary variables should be used in the model. Some guidelines on this problem have been given by the ESSnet Project (2012c). In the first place, detailed knowledge should be available of the nature of the source and concepts of the additional data such as the following (ESSnet, 2012c).

- Availability for all areas for which estimation required.
- Population scope of the data.
- Definitions of variables / concepts used.
- Purpose of data collection.
- Reference period – concurrent with survey period is ideal but often not available e.g., Census data may be from several years earlier – this is usually acceptable.
- Methodology of data collection.
- Survey design used if data is from sample survey.

- Quality of framework used for unit selection.
- Extent of missing data and whether any imputation used.
- Classifications used.
- Editing or data validation process used.

Additionally, auxiliary data should not contain missing values, as these may cause model bias and area level variables should not contain too many zero values as these may have a negative effect on model fit.

Also, at the initial stage it is important to include in the model, variables which are theoretically known to affect the variable of interest. The theoretical relationship must come from tested social or economic theories. Careful consideration should be given to understanding any significant differences between auxiliary data and variables of interest (Kordos, 2016). Then the correlation of these variables with the target variable should be checked. A strong relationship between the auxiliary data and the population of interest allows easy identification of "good" auxiliary variables. Potential relationships can be analyzed through scatter plots, correlations, or simple models.

In common with many other modelling fields, although the initial set of variables may be large, the final set should be relatively small. The optimal set of variables should be chosen in such a way as to give a model with the highest possible explanatory power (Claeskens and Hjort, 2008). Reducing the initial set can be based on a mixture of data investigative and automated statistical procedures. The following procedure is suggested in the guidelines given by the ESSnet project (2012c). First, correlations between pairs of auxiliary variables should be checked. If the correlation is greater than 0.9, then it should be ensured that both variables will not be included together in the final estimation model[11]. This is done to avoid the instability associated with near multicollinearity[12] of explanatory variables. Then from the remaining variables it would be right to pay special attention to those with the highest correlation

[11] This process is not binding as sometimes some of the strongly correlated pairs of variables when combined with other variables can give a model with very high explanatory power. In this case such a pair could eventually be included in the final model

[12] Multicollinearity refers to the linear relation among two or more variables. It can exist between two variables or between one variable and a linear combination of others. High correlation (correlation is a special case of multicollinearity, is the linear relationship between only two variables) implies multicollinearity. Multicollinearity is a data problem which may cause serious difficulty with the reliability of the estimates of the model parameters. For instance, if there is multicollinearity the signs of the regression coefficients in a multiple linear regression model may be wrong, i.e., different from the signs of correlation between the corresponding explanatory variable and the response variable (Alin, 2010)

with the variable of interest and to eliminate those with little or no correlation with the variable of interest. Also, it is possible that a transformation of a variable will give better explanatory power.

Another approach used in the ESSnet project (2012a), in the cases developed by France, Poland, and The Netherlands, is a step-forward procedure. In this case the process starts with models consisting of individual variables. Variables can then be added gradually as the model improves in terms of a selection criterion. If the software allows, automated procedures can be used so that this proccess can be completed in much less time.

Furthermore, in order to reduce the dimension of the covariate space, Principal Component Analysis (PCA) (Hastie, Tibshirani and Friedman, 2003) was applied in the ESSnet project (2012d) in the case developed by The Netherlands. PCA is a linear transformation of the original covariate space into a space where the dimensions are orthogonal directions of maximal variance. The coordinates of this space are referred to as the principal components. The principal components that correlate sufficiently well with the target variables are selected in the model. As the principal components are linear combinations of the original covariates, all original variables may contribute to the principal components which are retained in the model and might result in models with a substantially reduced number of covariates.

Furthermore, to select a suitable set of covariates in regression models, Kuo and Mallick (1998) propose a method based on expanding the regression equation to include all possible subsets of covariates so the vector of indicator variables dictates which set of predictors to include. They follow a Bayesian approach and obtain the posterior distribution of the indicator vector through the Markov Chain Monte Carlo (MCMC) method. The subset with the highest posterior probability gets selected. Also, Tibshirani (1996) proposed another method of selecting variables called LASSO (Least Absolute Shrinkage and Selection Operator). LASSO is a method that applies shrinkage factors to regression coefficients, and thus can more efficiently perform stable covariate selection. The procedure can select a few covariates that are related to the dependent variable from a large number of possible covariates. LASSO-based methods use "penalized regression" models that impose constraints on the estimated coefficients that tend to shrink the magnitude of the regression coefficients, often eliminating covariates entirely by shrinking their coefficients to zero. Therefore, nonzero coefficients are estimated for true covariates, whereas the coefficients for irrelevant variables are zeroed out.

Finally, it should be noted that when modeling is used, some model specific errors are introduced. For example, significant measurement error in the observed auxiliary variables may not be correctly accounted for by the model. To overcome such problems, various investigatory procedures are performed, known in the literature as model diagnostics. These procedures, which will be described in paragraph 2.6, determine the ability of the selected model to estimate the variable of interest. So, a model which does not pass on relevant diagnostics should be discarded. The final model should have good explanatory power and include a relatively small set of auxiliary variables.

2.6 Model selection and model diagnostics

Model selection and validation play a vital role in SAE methods. According to Rao and Wu (2001), if the assumed models do not provide a good fit to the data, the model-based estimators will be model biased which in turn can lead to erroneous inferences (inferences from model-based estimators refer to the distribution implied by the assumed model). Therefore, the problem of small area estimation refers not only to the production of reliable estimates but also on how to assess the estimation error, to evaluate the quality and plausibility of the estimates. In numerous cases, this quality assurance process constitutes the most complex part of the estimation procedure.

In order to compare different models and evaluate their performance as well as to check whether a small area model was producing adequate estimates, several diagnostic tests were developed and employed. Generally speaking, the process of selecting the 'best' model-based estimator involves the following three steps (ESSnet, 2012c):

- Model selection from a set of plausible models.
- Model fitting to adjust the selected model.
- Model diagnostic to check the adjusted model.

The above procedure can be repeated and continued so that we come to a satisfactory model.

2.6.1 Model selection. Once we have selected a set of variables (as described in paragraph 1.5), we can apply several criteria to select satisfactory models. According to Müller, Scealy, and Welsh (2013), the problem of model selection in a linear mixed model is complicated because selection of the covariance structure is not straightforward due to computational issues and boundary problems arising from positive semidefinite constraints on covariance matrices. Two basic model selection approaches in linear mixed model are information criteria such as AIC (Akaike, 1974) or BIC (Schwartz, 1978) and the Fence procedure (Jiang, Rao, Gu, and Nguyen (2008)).

Information Criteria. Information criteria are applied in practice by finding the model that minimizes an estimate of a criterion. Some basic information criteria are Akaike information criterion (AIC) (Akaike, 1974), Bayesian information criterion (BIC) (Schwartz, 1978) and conditional Akaike Information Critieria (cAIC) (Vaida and Blanchard, 2005).

The AIC criterion is based on the Restricted Maximum Likelihood (REML) method and is given by the formula (ESSnet, 2012a):

$$AIC = -2l + 2p \tag{2.6.1}$$

where, l is the log-likelihood at the parameter estimates and p is the number of parameters estimated in the model. The AIC compares models with the same fixed effects but different random effects. Since AIC is an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, a lower AIC means a model is considered to be closer to the truth.

The BIC criterion is also based on REML and is given by the formula (ESSnet, 2012a):

$$BIC = -2l + log(n)p \tag{2.6.2}$$

where, n is the number of the observations (areas in the Fay-Herriot model). The BIC permits comparison of models with different random effects, unlike the AIC. BIC is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is considered to be closer to the truth. Note that the BIC penalizes models with a greater number of variance parameters more than the AIC does. As a result, the two criteria may lead to different models.

In a parametric regression context Yang (2005) shows that there is a conflict between AIC and BIC: AIC is minimax-rate optimal but not consistent, BIC is consistent but not minimax-rate optimal. Also, the author points out that an adaptive model selection may solve this conflict, in which, in contrast to AIC and BIC, the penalty term[13] is data dependent. In addition, the AIC information criterion is found to be unsatisfactory for linear mixed models (Han, 2013). In this context, a proposed model selection that is well-suited for small area estimation is conditional AIC (cAIC) (Vaida and Blanchard, 2005). This model is relevant to inferences regarding the clusters, or areas, in the context of linear mixed models. The cAIC is given by the formula:

$$cAIC = -2\,log\,p\,(y|\boldsymbol{\beta}, \boldsymbol{v}) + 2p_{eff} \tag{2.6.3}$$

where, $p(y|\boldsymbol{\beta}, \boldsymbol{v})$ is the conditional likelihood for fixed and random effects, vectors $\boldsymbol{\beta}$ and $\boldsymbol{v}$ evaluated at their estimated values and y is the data. In the cAIC, the penalty for the model complexity is the effective degrees of freedom of the mixed model and is defined as the trace of the hat matrix, which maps the observed data to the fitted values, (Hodges and Sargent, 2001). As with AIC and BIC, so with cAIC the lower it is the closer to the truth the model is considered.

Either maximum likelihood (ML) or restricted maximum likelihood (REML) estimates for the variance components can be used in $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{v}}$. Conditional AIC is more relevant than AIC when the focus is on estimation of the realized random effects $\boldsymbol{v}$ and the regression parameters $\boldsymbol{\beta}$ (small area estimation falls into this category) (Rao, 2015). Furthermore, Han (2013) studied cAIC for the Fay–Herriot area level model given by (2.3.8). The proposed cAIC depends on the method of estimating σ_v^2, the variance of the random effect v_i in the model. Also, it performs better than the simple cAIC, given by (2.6.3), which ignores the error in estimating σ_v^2 and it is a suitable criterion measure with easy implementation for the Fay–Herriot model.

Fence procedure. Fence procedures are a class of strategies introduced for mixed model selection by Jiang et al. (2008), which includes linear and generalized linear mixed models. The basic idea behind fence methods consists of two steps: (1) Isolate a subgroup of "correct" models by constructing a statistical fence that eliminates incorrect models. (2) Select the "optimal" model from the subgroup according to a suitable criterion. If the minimal dimension criterion is used as optimality criterion, then the implementation of the fence method can be simplified by checking each candidate model, from the simplest to most complex, and stopping the search once a

[13] The final terms in both formulas 2.6.1 and 2.6.2 represent a penalty on the log-likelihood as a function of the number of parameters *p*.

model that falls within the fence is identified, and other models of the same dimension are checked to see if they belong to the fence. The Fence method is computationally very demanding, particularly because it involves the estimation of the standard deviation of the difference of lack-of-fit measures, for example, the negative log-likelihood.

Other model selection approaches in linear mixed model are shrinkage methods based on penalized loss functions such as LASSO, and Bayesian techniques (Müller et al. 2013).

2.6.2 Model diagnostics. In order to compare different models and evaluate their performance as well as to check whether a small area model was producing adequate estimates several diagnostic tests have been proposed. Diagnostic tests have been developed and employed for assessing the fit of a set of model-based small area estimates and performance of estimation methods as well as the plausibility of the outputs.

"Once one or several models have been selected, it is necessary to assess the fitting quality of the models. Different tools can be applied such as residual analysis to check if the model assumptions are fulfilled" (ESSnet, 2012c, p. 53). As noted in ESSnet (2012a) in the case study developed by ONS (2004, p. 85):

> Residual analysis is a way of checking that the model assumptions are satisfied, and the model accurately describes the population. In this case, two things are tested: model misspecification and nonconstant variance of the residuals (heteroscedasticity). If any pattern remains in the residuals this implies model misspecification e.g., a covariate influential to predicting the response variable has been left out of the model. It is also important to check the assumption of constant variance in the area level residuals since this will have an impact on the calculation of the confidence intervals.

Some tools used for residual analysis are:

- Histograms, Q-Q plots and boxplots to study the distribution of the selected models.
- Plots of model-based estimates versus residuals to reveal patterns and to identify residuals.
- Maps of residuals to check for spatial randomness in the case of geographic small areas.

Furthermore, Brown, Chambers, Heady and Heasman (2001) proposed four diagnostics to test the accuracy, validity and consistency of the small area estimates.

These diagnostics are based on the comparison among model- based estimates and direct estimates and include, the bias diagnostic, the coverage diagnostic, the goodness of fit diagnostic and the calibration diagnostic.

Bias diagnostic. A common diagnostic tool to check for overall bias is based on fitting a regression line to the scatter plot of the indirect estimates vs. direct estimates. This diagnostic tool provides both a visual illustration of bias and a parametric significance test (ESSnet, 2012b, p. 107). On the one hand the SAE estimates (X-axis) are plotted on a cartesian plane against the direct estimates (Y-axis) to verify if there is a deviation of the regression line between model based and direct estimates from y=x. From a visual point of view, careful examination of the plot could reveal systematic patterns in the data, which may require further investigation. On the other hand, a parametric test for the slope and for the intercept is carried out to check the unbiasedness of the model predictions. Considering that model-based estimates should be unbiased predictors of the direct estimates, it is expected that the slope of the regression line will not be significantly different from 1 and a very small intercept term, not significantly different from 0. When there is significant variation in small areas sizes this test requires an initial transformation of both the direct and model-based estimates, in order to satisfy the homoskedasticity assumption underpinning the fitting method (Brown et al., 2001).

Coverage diagnostic. Coverage diagnostic is a non-parametric significance test of the bias of model-based estimates relative to their precision, in order to evaluate the validity of the confidence intervals generated by the model-based estimation procedure (Ambler, Caplan, Chambers, Kovacevic and Wang, 2001). It assumes that direct estimates can generate valid 95% confidence intervals for the small area values of interest. Then the basic idea is to measure the overlap between the confidence intervals of the direct and the model-based estimates (across areas) and compare to the Binomial distribution (H0: p=0.95). However, because the degree of overlap between two independent 95% confidence intervals for the same quantity will be higher than 95%, it is necessary to modify the nominal coverage levels of the confidence intervals that are being compared to ensure a nominal 95% overlap.

Goodness of fit diagnostic. Goodness of fit diagnostic examines whether the model estimates are close to the direct estimates when the direct estimates are good (Brown et al. (2001)). To evaluate this, the squared differences between direct and model-based estimates inversely weighted by their variances, are calculated. Then these differences summed over all the domains. This sum gives more weight to differences from good direct estimates than from bad and is tested against the χ^2 distribution to provide a parametric significance test of bias of model estimates relative to their precision. Finally, results are provided using a Wald goodness of fit statistic.

Calibration diagnostic. This measure is based on what is typically a key requirement for small area estimates -that they sum to direct estimates at appropriate levels of aggregation. This property is known as calibration (Ambler et al., 2001). By calculating the relative difference between the aggregated model-based estimates prior to this calibration and the aggregated direct estimates, a measure of how accurate are the aggregated model-based estimates is obtained which provides a means to compare different models. Therefore, by computing this diagnostic, we obtain an accurate measure of the calibration property of the model estimates, providing also evidence of the presence/absence of spatial bias/autocorrelation. Finally, one issue to keep in mind is that when using this diagnostic, it is important to determine the appropriate calibration level.

In addition, some other diagnostic tools widely used to check the quality of a small area estimator are Coefficient of Variation (CV), Mean Squared Error estimation (MSE), MAPS and the Monte-Carlo simulation.

Mean Squared Error estimation. As mentioned in ESSnet (2012a, p. 26) "one of the most relevant measures of the quality of an estimator is the Mean Squared Error (MSE)". The mean squared error (MSE) is the most common measure to assess the uncertainty associated with the area-specific prediction under the model that has been assumed (Tzavidis, Zhang, Luna, Schmid and Rojas-Perilla, 2018). It is therefore important that any small area estimate is accompanied by an estimate of its MSE. The calculation of MSE is done by analytic or resampling methods. The model-based small area estimates should have MSE significantly lower than the variances of corresponding direct estimates. While EBLUP is relatively easy to obtain, estimation of its MSE is a challenging problem. In the context of model-based estimators, since the inference concerns the underlying model, the MSE estimation process is performed in parallel with the creation of the optimal prediction for which the error should be estimated. For instance, an approach to the MSE of the EBLUP can be achieved by summing up components that reflect uncertainty due to model adjustment, the ignorance of fixed effects, and the estimation of variance components. The unbiased estimation of each component depends on the application of the standard variance component estimators

(REML, ML or method of moments) (ESSnet, 2012a). Rao (2003) gives an appraisal of inferential issues in small area estimation focusing on the developments to estimate the MSE of model–based estimators. More specifically, Rao develop the outline of available methods to estimate the MSE of model-based estimators (analytic and resampling ones) providing a link between the methods and highlighting the rationale for the use of different estimators. Developments in confidence interval estimation under a basic area level model are also presented. Also, Jiang and Lahiri (2006) present an analysis of model-based small area estimators and the estimate of the MSE and a summary of new developments is included in ESSnet Project (2012a).

Coefficient of Variation. As with MSE, coefficient of variation (CV) can also be used as a measure of quality. The model-based small area estimates should have CV significantly lower than the CV of corresponding direct estimates (Molina and Morales, 2009; Molina and Rao, 2010). As mentioned in ONS (2004), estimates are considered precise and are suitable for publication when the majority of coefficients of variation are below 20%. Also, Molina and Marhuenda (2015) pointed out that “national statistical institutes are committed to publish statistical figures with a minimum level of reliability. A generally accepted rule is that an estimate with CV over 20% cannot be published” (p.86).

MAPS. Map production, in coloring scale, of small areas estimates can be a useful tool for validating results, externally, if small areas are geographical. This is due to the fact that the user can detect unexpected spatial patterns whose correct interpretation can lead to further improvements in the model (ESSnet, 2012a). Most studies done in SAE are accompanied by such maps (Molina and Morales, 2009; Molina and Rao, 2010; Szymkowiak, Młodak and Wawrowski, 2017).

Monte Carlo simulation. Monte Carlo simulation is an empirical way to carry out the comparative assessment. It is a particularly useful tool to evaluate the performance of the estimators that are generated as part of the estimation process. In Small Area Estimation, a common reference about Monte Carlo simulations is the EURAREA project (2004). Generally speaking, these simulations comprise the following steps.

- An artificial population is constructed, in which the true values of each component unit are known.

- A sample is drawn from this population base. This sample should reproduce as well as possible the real-life sampling scheme.
- Small area estimation procedures are applied to the data samples selected in the previous step.
- Second and third steps are repeated many times with independent applications.
- The area-level estimates generated by the samples are compared with the true values from the population for the same areas and summary performance statistics are computed.

2.7 Small Area Estimation in practice

The importance and necessity of SAE methods is demonstrated by the fact that in recent decades more and more national statistical institutes and other organizations around the world have been using small area statistics to make policy decisions. Some of the most important SAE projects around the world are presented below.

2.7.1 EURAREA project. The EURAREA project[14] (Enhancing Small Area Estimation Techniques to meet European needs) was a research programme funded by Eurostat within the Fifth Framework Programme of the European Union from 2000-2004 (EURAREA (2004)). The project was implemented with the participation of teams from seven European countries and spread across twelve National Statistics Institutes and Universities. The project was coordinated by the UK Office for National Statistics (ONS). The countries that took part were Spain, United Kingdom, Italy, Sweden, Norway, Finland and Poland.

The purpose of the EURAREA project was to provide European statisticians, particularly government statisticians, with the information they needed to decide when and how to use SAE methods in the production of official statistics.

All programs were released with open codes written in SAS language. The SAS program code was developed by the ONS for estimating small area means, their mean square errors and confidence intervals. In particular, the estimators calculated were:

- The national sample means
- The direct estimator
- The GREG estimator with a standard linear regression model

[14] http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/spatial-analysis-and-modelling/eurarea/index.html

- The synthetic estimator considered under two different models: a linear two-level model with individual data and a linear model with area-level covariates
- The EBLUP estimator using two different models: a linear two-level model with individual data and a linear model with area-level covariates

The analysis in each country was performed for two levels of the European Nomenclature of Territorial Units for Statistics (NUTS[15]) hierarchy. In all six countries estimates were produced for NUTS3 areas and for an area smaller than NUTS3. Also, real data from population Censuses or population registers were used to create a simulation database. Repeated samples were drawn from this database and one or more small area estimation methods were applied to each sample. The simulation settings make it possible to compare the area-level estimates, which were generated by the samples, with the true values from the population for the same areas.

The same target variables were used in the simulations for each country. The variables examined were:

- Average equivalized household net income.
- The proportion of single person households.
- The proportion of the economically active population who are unemployed according to International Labour Organisation (ILO)

The project results showed that at NUTS3 level model-based estimators achieved comparable or slightly better levels of precision than design-based estimators. At areas smaller than NUTS3 level, model-based estimators substantially outperformed design-based methods. According to the results of the project, Eurostat encourage member states to adopt model-based methods of estimation for area sizes below (and possibly including) NUTS3.

2.7.2 ESSnet on Small Area Estimation project. An ESSnet project is a network of several European Statistical System (ESS) organizations aimed at providing results that will be beneficial to the whole ESS[16]. One of these projects was ESSnet on Small Area Estimation [17] . It was implemented from December 2009 to March 2012 and was partly financed by Eurostat. The countries that participated in the project were

[15] See more details about NUTS in paragraph 4.2.3
[16] https://ec.europa.eu/eurostat/cros/content/essnet-generalities_en
[17] https://ec.europa.eu/eurostat/cros/content/sae-finished_en

France, Germany, Italy, Netherlands, Norway, Poland, Spain, Switzerland and the United Kingdom.

The main purpose of this project was to develop a framework in order to product small area estimates for ESS social surveys. The objective was to produce a set of principles that point out some good practices to follow in order to obtain step by step small area estimates of interest with a sufficient level of quality. Various SAE fields of survey have been examined (not necessarily the same in every country) such as:

- Analphabetism
- Labour Force Survey (LFS)
- Health Survey (HS)
- Consumer Expenditure Survey (CES)
- Crime and Victimization Survey
- Structural Business Survey
- Poverty Indicators

The software used for SAE applications in the NSIs, involved in the ESSnet SAE project, was SAS and R. Most of the applications are carried out using SAS software and a small part of the studies makes use of R. The software routines specifically developed in the ESSnet SAE project are all R functions that can be run with general data sets. All codes written are open.

The results of the project were as follow:

- The available documents on small area estimation were updated and a common knowledge created on application of small area estimation methods
- Suitable criteria were reviewed and developed to assess the quality of SAE methods for the choice of proper model and the evaluation of MSE.
- Practical guidelines were provided in the context of ESS social surveys.

2.7.3 BIAS project. BIAS (Bayesian methods for combining multiple Individual and Aggregate data Sources in observational studies)[18] is a project based at Imperial College in London. It sits within the Economic and Social Research Council's National Centre for Research Methods framework. BIAS I was funded between April 2005 and June 2008 under the first phase of node commissioning. BIAS II was funded by the second commissioning phase from July 2008 to June 2011. The objective of the

[18] www.bias-project.org.uk/

project was to provide researchers with guidelines on how to combine data from multiple sources and improve existing social science methods to manage complexity of observational data. To achieve the above goals, Bayesian hierarchical models were used as they provide a natural way for linking together many different sub-models and data sources.

The BIAS I research programme consisted of three methodological components. One of them was the Small Area Assessment and was carried out in collaboration with ONS. The main objective was to estimate in each small area some indicators such as income, crime rate, unemployment, etc., using individual-level data from various surveys as well as area-level auxiliary variables, from Census and administrative sources. Existing estimation methods were used by ONS to incorporate spatial and spatio-temporal dependence, and to compare likelihood and Bayesian methods for small area estimation.

All methods and techniques of small area estimation considered in the BIAS project were implemented in R and WinBUGS and were described in materials available on the project web page. Two of the functions that have been released were EBLUP.area and SEBLUP.area. These functions allow users to compute small area estimates using EBLUP estimators based on area level models.

2.7.4 SAMPLE project. The S.A.M.P.L.E. (Small Area Methods for Poverty and Living Condition Estimates) project[19] was a research programme funded by the European Commission under the Seventh Framework (FP7) Programme of the European Union. The project started in March 2008 and ended in March 2011 (SAMPLE, 2009).

The main purpose of the project was to identify and develop new indicators and models that will help the understanding of inequality and poverty within the small area estimation framework. Particular attention was paid to social exclusion and deprivation. The data used to achieve the above purpose were combined by national surveys with data from local administrative databases.

SAS and R software was given for applying indicators (on EU-SILC) proposed by the project. Different R functions were provided for fitting:

- area level models,

[19] www.sample-project.eu/

- spatially correlated random effects area level models,
- temporal models with independent and correlated temporal effects,
- partitioned area level time models,
- unit-level models with independent and correlated time effects,
- spatial Fay-Herriot model with uncorrelated and correlated time effects.

The methodology developed by the consortium could be applied to other areas of social research. For example, the small area techniques that were used can be employed in estimating educational inequalities at small areas. Also, through this project a step was taken for a better understanding of record linkage processes required for integrating administrative with survey data and the problems associated with administrative data such as self-selection bias.

2.7.5 AMELI project. The A.M.E.L.I (Advanced Methodology for European Laeken Indicators) project[20] was a research programme funded by European Commission under the Seventh Framework (FP7) Programme of the European Union. The project started in April 2008 and ended in April 2011.

The main target of this project was to develop an improved methodology in order to measure social cohesion adequately with Laeken indicators while regarding national characteristics and practical peculiarities from the EU-SILC. The study included research on data quality including its measurement, treatment of outliers and nonresponse, small area estimation and the measurement of development over time. Also, a simulation study was performed based on EU-SILC data and a software framework was developed using R.

2.7.6 SAIPE project. The S.A.I.P.E (Small Area Income and Poverty Estimates) project[21] is a program running from the U.S. Census Bureau and provides annual estimates of income and poverty statistics for all school districts, counties, and states. SAIPE data also produces single-year poverty estimates for the school-age population (age 5-17) for all school districts in the U.S.

The main purpose of this program is to provide estimates of income and poverty so that federal programs can be properly managed and the optimal allocation of the corresponding funds to local jurisdictions can be achieved. The data they use combine survey data with population estimates and administrative records. Due to the

[20] https://ameli.surveystatistics.net/
[21] https://www.Census.gov/programs-surveys/saipe.htm

comprehensive geographic coverage and one-year focus, SAIPE data can be used to analyze geographic variation in poverty and income, as well as changes over time. The county and state models both follow the general form suggested by Fay and Herriot (1979).

Other examples of major SAE programs in the United States include the following:

- The Bureau of Labor Statistics' Local Area Unemployment Statistics (LAUS) program that produces monthly and annual estimates of employment and unemployment for states, metropolitan areas counties, and certain subcounty areas.
- The National Agricultural Statistics Service's County Estimates program that produces county estimates of crop yield (USDA 2007).
- The estimates of substance abuse in states and metropolitan areas that are produced by the Substance Abuse and Mental Health Services Administration.

3. Estimating Poverty Using Small Area Estimation Methods

3.1 Introduction

In recent decades, the fight against poverty has become increasingly central, and recent radical economic and social transformations have sparked new interests in this area. Levels of poverty are used first and foremost for monitoring social and economic conditions and for providing benchmarks of progress or failure (Pratesi, 2016). Therefore, it represents not only a problem but also a symptom of the ineffectiveness of policies to improve living conditions. Fighting poverty is one of the goals of the European Commission, which is clearly emphasized in the "Europe 2020" strategy[22]. In its 2014–2020 multiannual financial framework the EU has budgeted one trillion euros to support growth and jobs and reduce poverty and social exclusion. Success depends on developing the right policies and programs and targeting them effectively. Applying appropriate social policy, however, requires adequate knowledge of the extent of the poverty phenomenon. This knowledge depends both on the definition of a set of comparable and readable poverty indicators and on the data provided on poverty and living conditions. Such information is provided through surveys of living conditions conducted, among others, by the Central Statistical Office (CSO) of each country. However, the sample size in these surveys allows for an accurate estimate of the poverty rate only at a very general level such as the whole country and regions. Estimating poverty at a lower level of spatial aggregation is vital as it allows governments to formulate and target policies and thus allocate funds to small areas as well (Rao, 2003, p.3). The solution to the problem requires a more sophisticated statistical approach such as small area estimation methods. Within the small area estimation framework major projects such as SAMPLE, AMELI and SAIPE[23] have been carried out in order to measure and understand poverty. For instance, the SAIPE project uses a Fay-Herriot area level model in order to produce model-based county estimates of school-age children under poverty. Also, the World Bank (WB) has been

[22] In June 2010, the European Council adopted the Europe 2020 Strategy which is the EU's growth strategy for the current decade, aiming at developing in the EU a smart, sustainable and inclusive economy. In this context, the European Council adopted a social inclusion target, namely raising at least 20 million people from the risk of poverty and exclusion by 2020 (European Commission (2010)).
[23] More information about these programs is given in the paragraphs 2.7.4, 2.7.5, 2.7.6.

releasing small area poverty and income inequality estimates for some countries, using the methodology of Elbers, J.O. Lanjouw, and P. Lanjouw (2003)[24].

3.2 Poverty measurement

Poverty, as many other socio-economic phenomena can be measured, but it is not an easy task. According to Sen (1976), the specific way of measuring a phenomenon should depend on the purpose for which the measure will be used. In the case of poverty measurement, there are various purposes and applications for the resulting measures: Poverty may be measured by a government to provide a continuous assessment of how its various policies are affecting the conditions of the poor, or it can be measured to help uncover the causes and correlates of poverty in order to formulate policies to combat it. In addition to the above purposes, a standard use of the poverty methodology is to enable governments to identify individuals or families who are in poverty and, therefore, to focus services and policies directly on them.

As pointed out by Decancq, Goedemé, Van den Bosch, and Vanhille (2013) to measure poverty, three key issues need to be addressed. First, the appropriate metric of individual well-being must be determined. Second, a cut-off value must be set or a threshold below which individuals are considered poor and third, it is necessary to select an aggregation procedure (poverty indicators) to attain a poverty rate for society as a whole.

3.2.1 Definitions of poverty. The first step in measuring poverty is to formulate an appropriate definition. Clarification of how poverty is defined is extremely important as different definitions involve the use of different indicators for measurement. This can lead to the identification of different individuals and groups as poor and therefore the implementation of different policies to reduce poverty. The discussion over a definition of poverty has lasted for many years. There is no single definition of poverty and it has been characterized as either "having less than an objectively defined minimum, having less than others in society or feeling that you do not have enough to get along" (Hagenaars and De Vos, 1998). A European definition was first agreed by the European Council in 1975 and sees poverty as "individuals or families whose resources are so small as to exclude them from the minimum acceptable

[24] Elbers et al. (2003) assumed a unit level model that combines both Census and survey data. Using that model, they produce disaggregated maps that describe the spatial distribution of poverty and inequality.

way of life of the Member State in which they live" (Council Decision 75/458/EEC) and on December 1984, the European Commission extended the definition as: "the poor shall be taken to mean persons, families and groups of persons whose resources (material, cultural and social) are so limited as to exclude them from the minimum acceptable way of life in the Member State in which they live" (EEC, 1985). The above definition is linked with many approaches to measuring poverty and consists of two basic concepts: the concept of resources and the concept of the least acceptable way of life. Resources are referred to in the broadest sense, not only cash, wealth and other income, but also human resources such as health and education. The concept of the least acceptable way of life is related to Sen's view of basic 'functionings' or 'capabilities' (Sen, 1983, 1985). Functionings are the doings (such as having a good job, having a decent standard of living etc.) and beings of individuals (such as being safe and healthy, being able to appear in public without shame etc.). Capabilities are the set of potential functionings that a person can acquire. Many authors, such as Townsend (1979), have suggested similar definitions.

3.2.2 Measurement approaches. Once a proper definition of poverty has been formulated, the next step is to translate that definition into a computable measure of poverty. In practice, a wide variety of poverty measures are used, some are simple, and others are complex. Poverty measurement approaches can generally be classified into one-dimensional and multidimensional (SAMPLE, 2009). The one-dimensional approach is the traditional poverty approach (or monetary approach) and is characterized by a simple dichotomization of the population into poor and non-poor defined in relation to some chosen poverty line that represents a percentage (generally 50%, 60% or 70%) of the mean or the median of the equivalent income distribution. This approach refers to only one proxy of poverty, namely low income or consumption expenditure and it takes place in two different and successive steps. The first aims to identify who is poor and who is not, depending on whether a person's income is below or a above a critical threshold, the so-called poverty line. The second step is to summarize the amount of poverty in aggregate indices determined in relation to the income of the poor and the poverty line.

Nowadays the multidimensional approach is gaining more and more ground. According to Stiglitz, Sen and Fitoussi (2009) a multidimensional approach to poverty is needed to define the concept of "well-being". Also, Narayan (2000), based on a large-scale survey, states that the poor worldwide perceive "well-being" and poverty as multidimensional concepts. The basic concept of the multidimensional approach is that poverty as a complex phenomenon cannot be limited to the monetary dimension but must also be explained by other variables whose effects on poverty are not covered by income. Enjoying good health, gaining access to an appropriate education and training system, having a good job, living in a context of social relations based on trust and environmental care are relevant dimensions of well-being. Therefore, in order to measure poverty, it is necessary to expand to a variety of non-monetary indicators of living conditions and at the same time to adopt mathematical tools that can represent the complexity of the phenomenon. The multidimensional nature of poverty is a widely recognized fact, not only by the scientific community, but also by many official statistical agencies and international institutions. For instance, in December 2001, the Laeken European Council endorsed a set of statistical indicators to highlight the multidimensional nature of poverty.

Both monetary and multidimensional poverty measures are valuable in identifying the poor and formulating appropriate policies. They provide different knowledge and complement each other. As mentioned in UNECE (2017, p. 18) “what can be surprising is the common finding that people who are multidimensionally poor, or deprived in non-monetary indicators, are not necessarily income poor. Divergences between monetary poverty and multidimensional poverty indicators mean that both need to be measured”.

3.2.3 Choosing poverty lines. Suppose a measure of household well-being is chosen, such as consumption expenditure. The next step is to choose a poverty line. Households whose consumption expenditure falls below this line are considered poor. There are three basic approaches to establishing a poverty line: “The absolute poverty line (or “having less than an objectively defined absolute minimum”), (Hagenaars and De Vos, 1988), the relative poverty line (or “having less than others”), and the subjective poverty line (or “feeling you do not have enough to get by”), (UNECE, 2017, p. 63).

An absolute poverty line is a fixed cut-off level applied across all potential income distributions, after adjusting for differences in purchasing power. Compared over time, the line is unchanged (except for adjustments for changes in price levels) even in view of economic growth. To obtain the absolute poverty line various competing methods and assumptions are available, each of which can generate a different poverty cut-off. The most common approach to determining an absolute poverty line is to estimate the cost of a bundle of goods that is considered to meet basic consumption needs. Absolute poverty lines are used worldwide in both developed and developing countries and this is the approach used most to identify the poor over time and space. A concern that arises is how often the absolute poverty line should be updated. On the one hand it must be fixed enough to capture changes in poverty and on the other hand it must be updated frequently enough to reflect changes in economic circumstances. The United States poverty line has remained fixed since 1965 and the World Bank's main poverty standard was updated in 2015.

A relative poverty line begins with a definition of a standard of living for a given distribution of income (such as the mean, median or some quintile) and defines the cut-off as some percentages of this standard. The result is that the cut-off below which, one is considered poor, varies proportionally with its income standard. Relative poverty lines are most often used in developed countries where there is less concern about achieving a minimum absolute living standard and greater interest in inclusion or relative position (UNECE, 2017). One example is the European Union's country-level poverty lines, which are set at 60% of a country's median (disposable) income. A key advantage of a relative poverty line is its conceptual clarity and simplicity of use (particularly for international comparisons). However, any relative poverty threshold is essentially arbitrary, as the selection of living standard and the percentage of this standard could vary among countries according to social preferences.

Both the absolute poverty line and the relative one, are used frequently for poverty monitoring, but in both approaches some practical issues need to be addressed. An absolute poverty line could be set too low in developed countries while a relative poverty line could be set too high for developing countries. As a result, neither of these two concepts is satisfactory when one is computing poverty profiles of a heterogeneous set of countries (OECD, 2013). With the relative poverty line, the analysis of changes in poverty over space and time is less transparent. In contrast, with the absolute line, there are two sources of change: the direct impact of the change in the distribution and the indirect impact through the change in the underlying living standards, such as growth in median income. In order to overcome the above problems, alternative approaches have been proposed in the literature such as anchored poverty lines (OECD, 2013) and are both absolute and relative in the sense that a given relative line is computed for one period and then frozen and used as an absolute line over time. Another approach is the Hybrid poverty lines (Ravallion and Chen, 2011) that require poverty to fall when all incomes in a distribution rise by an equal proportion.

Furthermore, an alternative approach is the subjective poverty line which explicitly recognizes that poverty lines are inherently subjective judgments made by people about what constitutes a socially acceptable minimum standard of living in a particular society. This approach is often based on survey responses to a question such as the following: "What income level do you personally consider to be absolutely minimal?". In practice, few surveys include such subjective questions.

3.2.4 Poverty indicators. Having decided on a welfare measure and setting a poverty line, the next step is to select one or more indicators useful for combating poverty. Indicators can be used to indicate the level of poverty in different countries or regions, the depth of poverty people experience and how poverty changes over time. According to UNECE (2017) monetary poverty indicators can generally be classified into two categories of measures: static and dynamic measures.

Static measures are based on income or consumption at a given point in time. The most widely used static measure is the *headcount ratio* (also called poverty incidence) (SAMPLE, 2009). It measures the proportion of the population counted as poor, that is, the proportion living in households whose income or consumption expenditure is less than the poverty line. Suppose a population of size N and let z be a given poverty line. In addition, let E_i be the income/expenditure of the i-th population unit, with $i = 1,2,\dots,N$. Then the headcount ratio P_0 is given by (Molina and Rao, 2010):

$$P_0 = \frac{1}{N}\sum_{i=1}^{N} I(E_i < z) \tag{3.2.1}$$

$$\text{with, } I(E_i < z) = \begin{cases} 1, \text{ if } \mathrm{E}_i < z \text{ (population unit under poverty)} \\ 0, \text{ otherwise} \end{cases} \tag{3.3.2}$$

The headcount ratio is simple to construct and allows users to easily understand the scale of poverty between different groups, but it presents some weaknesses also. First, it violates the transfer principle of Pigou-Dalton[25], which states that transfers from a richer to a poorer person should improve the measure of well-being. The headcount ratio does not indicate the depth of poverty that people experience, and hence, does not change if people below the poverty line become poorer. In addition, it calculates the percentage of individuals and not households, as the estimates of poverty should be

[25] Pigou -Dalton principle originally suggested by Arthur Pigou (1912) and Hugh Dalton (1920)

calculated, making a not-always-true assumption that all members of the household enjoy the same level of well-being.

Another static measure of poverty is the *poverty gap index*. It measures the extent to which individuals fall below the poverty line and expresses it as a percentage of the poverty line (SAMPLE, 2009). The poverty gap index is given by (Molina and Rao, 2010):

$$P_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{z-E_i}{z}\; I(E_i < z) \tag{3.2.3}$$

where z is the poverty line and where E_i is the actual expenditure/income for the *i*-th population unit (poor people). Also, $G_i = \begin{cases} z - E_i, & \text{if } i\text{-th unit is poor} \\ 0, & otherwise \end{cases}$ is called the poverty gap. Equation (3.2.3) is the mean proportionate poverty gap in the population and shows how much would have to be transferred to the poor to bring their incomes or expenditures up to the poverty line. Dividing by the poverty line normalizes the measure so that comparisons can be made both between countries and over time. However, a disadvantage of the poverty gap indicator is that it cannot reflect changes in inequality between the poor as it only reflects the average depth of poverty. Also, another disadvantage is that when people leave poverty, it can increase rather than fall if the average poverty gap of those who remain increases.

In addition, a static measure that takes into account the inequality among the poor is the *squared poverty gap index* or *severity poverty index*, which is given by (Molina and Rao, 2010):

$$P_2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{z-E_i}{z}\right)^2\; I(E_i < z) \tag{3.2.4}$$

The squared poverty gap index is a weighted sum of poverty gaps where the weights are the proportionate poverty gaps themselves giving more weight to observations that fall well below the poverty line. Thus, takes into account inequality among the poor and emphasizes extreme poverty. However, the squaring of the poverty gaps, makes it difficult to interpret the results.

The measures described above belong to a family of measures proposed by Foster, Greer and Thorbecke (1984) called FGT poverty measures which may be written as:

$$P_a = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{z-E_i}{z}\right)^a\; I(E_i < z) \tag{3.2.5}$$

where the parameter a can be viewed as a measure of poverty aversion: A larger a gives greater emphasis to the poorest poor. As a becomes very large, P_a approaches a

"Rawlsian" measure which considers only the position of the poorest household. The case $a = 0$ coincides with the headcount ratio P_0, $a = 1$ gives the poverty gap index P_1 and $a = 2$ gives the squared poverty gap index P_2. FGT measures of poverty can be disaggregated for population sub-groups and the contribution of each subgroup to national poverty can be calculated (SAMLE, 2009).

As mentioned in UNECE (2017, p. 90) "despite the importance of tracking changes in the depth of poverty, measures such as the poverty gap index have had relatively limited use in policy formation and monitoring due to being deemed "unintuitive" and difficult to understand". A proposed measure that seeks to address this problem is the *person-equivalent approach*, developed by Castleman, Foster and Smith (2015).

> Person equivalent headcount measures benchmark the initial conditions of the poor, and then employ this standard as a measuring rod to count the number of standardized poor, or person equivalents. The picture of poverty is altered in appropriate ways: it raises the level of measured poverty when the conditions of the poor become worse; it lowers it when the average conditions are better (Castleman et al. 2015, p. 4)

In addition to the static measures mentioned above, there are also the so-called dynamic measures. While static measures describe current levels of poverty and how they vary over time, place and groups, dynamic measures use longitudinal data to examine poverty over time, as well as transitions to and from poverty. For instance, such measures are the *persistent poverty indicators* as well as the *entry and exit rates*. The *persistent poverty indicators* have been developed in the context of the generally accepted assumption that the longer an individual remains in poverty, the more detrimental it is. In particular, Fouarge and Layte (2005) have shown that the longer a person remains in poverty, the less likely they are to escape poverty. In addition, according to Dickesron and Popli (2014), persistent poverty adversely affects children's cognitive development, especially in the first years of life, and increases the likelihood of experiencing poverty as an adult. So, indicators that use longitudinal data are valuable to policymakers, as they can help identify the groups most likely to experience long periods of poverty. The most widely used persistent poverty indicator is the one used by the European Commission (2010). According to this indicator, the persistent at-risk-of-poverty rate is defined as the percentage of the population living in households where their equivalent disposable income is below a poverty line for both the current year and at least two of the previous three years. The calculation of this measure requires a longitudinal instrument, through which individuals are monitored for four years. Furthermore, another dynamic measure is the *entry and exit rates*. The entry rate is defined as the percentage of people who were not in poverty a year earlier but fell into poverty the following year. The exit rate is defined as the percentage of people who were at-risk-of-poverty the previous year but are not at-risk-of-poverty during the current year. Dynamic poverty measures are an important tool for effective policy development and targeting. As they examine transitions to and from poverty from one year to the next, this can be particularly useful where limited panel durations make it difficult to analyze the duration of poverty. As mentioned in UNECE (2017), National Statistical Institutes should look at opportunities to generate longitudinal data, either from surveys or from administrative sources, in order to produce comparable dynamic poverty indicators in the future.

3.3 The Laeken and AROPE indicators

The static and dynamic measures described above have advantages and disadvantages. For this reason, most countries and international organizations prefer not to focus on a single indicator, but to publish a series of indicators that give a more comprehensive picture of poverty. Also, it is essential that there are common indicators between countries in order to be comparable. In this context, the European Council has established a number of common indicators between its Member States such as Laeken and AROPE indicators. These indicators are regularly produced for every EU country on a comparable basis. EU-SILC and Labor Force Surveys are the main data sources used to calculate them.

In December 2001, at the Laeken European Council, EU Heads of State and Government endorsed a first set of 18 common statistical indicators of poverty and social exclusion (Laeken indicators), indicators that were refined later by the European Commission Social Protection Committee (2001). It was essentially a portfolio of indicators designed according to a set of methodological principles, as formulated by Atkinson, Cantillon, Marlier and Nolan (2002). This portfolio contained both indicators based on household incomes (monetary indicators) and indicators based on non-monetary symptoms of poverty (non-monetary indicators). It was organized in a two-level structure of 10 primary indicators[26] – covering the broad fields considered to be the most important elements leading to social exclusion – and 8 secondary indicators[27] – intended to support the lead indicators and describe other dimensions of the problem. In the following years, the portfolio was further expanded to include a wide range of indicators covering various aspects of social inclusion and social protection. The adoption of common indicators was carried out in the context of the so-called "open method of coordination" (OMC) which is a voluntary process for political cooperation based on agreeing common objectives and common indicators, which show how progress towards these goals can be measured. The development of European Statistics on poverty and social exclusion was inscribed in this framework as the means to ensure this measurement of progress towards common goals in social policy. The main objectives were comparability between Member States and the balance and transparency of the overall portfolio.

In June 2010, the European Council went a step further and set a specific goal in the Europe 2020 strategy: “20 million fewer people should be at risk of poverty and exclusion according to AROPE (At-Risk-Of-Poverty or social Exclusion) indicator” (European Council, 2010, p. 12). However, Member States, taking into account their national circumstances and priorities, were free to set their national targets using the appropriate indicators. The AROPE indicator is sourced from the EU Statistics on Income and Living Conditions (EU-SILC) and defines the share/number of people who are at risk-of-poverty or severely materially deprived or living in households with very low work intensity. Specifically, the AROPE consists of three sub-indicators that are derived from EU-SILC data (European Commission, 2001):

- a relative component: The *At-Risk-Of Poverty rate* / monetary poverty (AROP). This sub-indicator has been used in the European Union as the main indicator to monitor progress towards the eradication of poverty in the

[26] i) At-risk-of poverty rate + illustrative threshold values ii) Persistent at-risk of poverty rate iii) Relative median poverty risk gap iv) Long term unemployment rate v) Population living in jobless households vi) Early school leavers not in education or training vii) Employment gap of immigrants viii) Material deprivation (to be develop) ix) Housing x) Self-reported unmet need for medical care and Care utilization (European Commission Social Protection Committee (2001))

[27] i) At-risk-of poverty rate ii) Poverty risk by household type iii) Poverty risk by the work intensity of households iv) Poverty risk by most frequent activity status v) Poverty risk by accommodation tenure status vi) Dispersion around the at-risk-of-poverty threshold vii) Persons with low educational attainment viii) Low reading literacy performance of pupils (European Commission Social Protection Committee, 2001)

European Union until the adoption of Europe 2020. It is defined as the percentage of the population with an equivalized disposable income below the at-risk-of-poverty threshold, which in the EU is set in each country at 60 % of the national median[28] equivalized disposable income expressed in national currency (after social transfers). The equivalized disposable income of a household is defined as the total disposable income[29] of a household divided by the equivalized household size[30], defined according to the OECD-modified scale[31] . This scale gives a weight to all members of the household and then adds these up to arrive at the equivalized household size.

[28] The median income is the household income of what would be the middle individual if all individuals in the population were sorted from poorest to richest. As it represents the middle of the income distribution, the median household income provides a good indication of the standard of living of the "typical" individual in terms of income. The median is the most stable among other measures, and it is the most appropriate choice for a log-normal distribution (which often well approximates the distribution of income or consumption expenditure) as well as for predicting separately the effects of the economic cycle and inequality within the distribution. (UNECE, 2011, p. 73).

[29] **Table 3.3.1** Definition of Total Disposable Household Income

Total Disposable income = (**Total income**) – (**Current transfers paid**).
Current transfers paid: Direct taxes (net of refunds), Compulsory fees and fines, Current inter-household transfers paid, Employee and employers' social insurance contributions, Current transfers to non-profit institutions.
Total income = (**Current transfers received**) + (**Primary income**)
Current transfers received: Social security pensions / schemes, Pensions and other insurance benefits, Social assistance benefits (excluding social transfers in kind), Current transfers from non-profit institutions, Current transfers from other households.
Primary income = (**Property income**) + (**Income from employment**) + (**Income from household production of services for own consumption**)
Property income: Income from financial assets, net of expenses, Income from non-financial assets, net of expenses, Royalties
Income from employment: Employee income, Income from self-employment
Income from household production of services for own consumption: Net value of owner-occupied housing services, Value of unpaid domestic services, Value of services from household consumer durables

Note: Adapted from Table 2.1 in The Canberra Group (UNECE, 2011, p. 11)

[30] To take into account the impact of differences in household size and composition, the total disposable household income is "equivalized". The equivalized income attributed to each member of the household is calculated by dividing the total disposable income of the household by the equivalisation factor. Equivalisation factors can be determined in various ways.

[31] After having used the "old OECD scale" in the 1980s and the earlier 1990s, Eurostat adopted in the late 1990s the so-called "OECD-modified equivalence scale". This scale, first proposed by Haagenars, Vos, and Zaidi (1994), assigns a value of 1 to the household head, of 0.5 to the second and each subsequent person aged 14 and over and of 0.3 to each child less than 14 years.

- a "kind of" absolute component: *Material deprivation*. This sub-indicator includes people who suffer from severe material deprivation and have living conditions severely constrained by a lack of resources. They experience at least four out of the following nine deprivation items. They cannot afford i) to pay rent or utility bills, ii) to keep the house warm enough, iii) to face unexpected expenses, iv) to eat meat, fish or protein equivalent every other day, v) a week off away from home, vi) a washing machine, vii) a telephone, viii) a colour TV, or ix) a car.
- an exclusion from labour market component: *Severe low work intensity*. This sub-indicator includes people living in households with very low work intensity who are those aged 0-59 living in households where adults worked less than 20% of their total work potential during the past year.

The huge impact of the financial and economic crisis on increasing poverty in the EU has given an especially important visibility and policy relevance of the AROPE indicator.

3.4 The European Union Statistics on Income and Living Conditions

Poverty measurement depends largely on the availability and quality of appropriate data. Data may come from sample surveys at national level, meso databases at regional level, local databases, registers and sample surveys conducted at regional and local levels. In the EU there are several cross-national comparative surveys on the study of poverty and social exclusion. These include the survey on Ageing and Retirement in Europe (SHARE), the European Quality of Life Survey (EQLS), the European Social Survey (ESS) and the Survey of Health. However, these surveys either cover only a portion of the population such as SHARE or have a small sample size such as EQLS or contain only limited information on income and living conditions such as ESS. From 2004 onwards, the EU's main source for micro-data on income and living conditions is the annual EU-SILC (European Union Statistics on Income and Living Conditions) survey.

In 2004 EU-SILC replaced the European Community Household Panel (ECHP) as the common European source for data on income and social inclusion. ECHP ran as a long-term panel structure in fourteen European Member States over the eight-year period from 1994 to 2001. The EU-SILC is a yearly data collection effort conducted by

Eurostat in cooperation with the National Statistical Institutes (NSIs) of the European Union, European Free Trade Association (EFTA) and candidate countries. The first round of the EU-SILC was carried out in 2003 on the basis of a "gentlemen's agreement" in six Member States (Belgium, Denmark, Greece, Ireland, Luxembourg and Austria) and Norway (Eurostat 2016). The EU-SILC legal basis entered into force in 2004 and now covers all EU countries, Iceland, Norway, Switzerland, and some other countries participating on a voluntary basis. Since 2004 EU-SILC is conducted on the basis of EU legislature and microdata is made available free of cost to accredited researchers from 2004 onward in form of the EU-SILC User Database (UDB) (Eurostat 2015).

The primary objective of the EU-SILC is to provide comparable data on income, poverty, social exclusion and living conditions (Eurostat, 2013). Eurostat employs the EU-SILC as an important data source for indicators on income, poverty and living conditions in the EU within the Europe 2020 strategy, the EU's agenda for growth and jobs. EU-SILC pays additional attention to the sample design, internationally harmonized income definitions, and EU-wide coverage in order to overcome the quality problems that had arisen in the ECHP, such as the low response rates, steady attrition rates, incomplete geographical coverage and poor timeliness, (Clemenceau and Museux, 2007). All statistics under the Income and Living conditions (ILC) domain in the Eurostat dissemination database are EU-SILC data.

The EU-SILC mandatory variables cover basic information on the respondents' demographic traits, their involvement in the education process, the evaluation of health status, selected data on deprivation of basic necessities, data on housing conditions, detailed information on economic activity, and above all, an extensive range of information on the level and sources of income. Information on social exclusion and housing conditions are obtained for households. Education and health data are collected for persons aged 16 and over.

EU-SILC is implemented as a rotating panel study in which households are interviewed in four consecutive years and provides annually two types of data[32] (Eurostat, 2013):

[32] The sample for any year consists of 4 replications which have been in the survey for 1-4 years. With the exception of the first three years of survey, any particular replication remains in the survey for 4 years. Each year, one of the 4 replications from the previous year is dropped and a new one is added. In order to have a complete sample the first year of survey the four panels began simultaneously. For the

- Cross-sectional data concerning a given time or a specific period of time with variables on income, poverty, social exclusion and other living conditions, and
- Longitudinal data related to changes at the individual level over time, observed periodically for a period of four years.

Furthermore, in terms of the units involved, four types of data are involved in EU-SILC (Eurostat, 2013):

- variables measured at the household level.
- information on household size and composition and basic characteristics of household members.
- income and other more complex variables termed 'basic variables' (education, basic labour information and second job) measured at the personal level, but normally aggregated to construct household-level variables; and
- variables collected and analyzed at the person-level 'the detailed variables' (health, access to health care, detailed labour information, activity history and calendar of activities).

The reference population of EU-SILC is all private households and their current members residing in the territory of the Member States at the time of data collection. One characteristic of EU-SILC is flexibility in terms of data sources and sampling design. Eurostat strongly encourages the use of existing data sources, whether they are surveys or registers and the use of national sampling design. While the basic rules (such as definitions, minimum effective sample size, time reporting, etc.) are legally binding and therefore common to all participating countries, there are significant differences in terms of sample design and collection and processing of the data. These differences often make comparability of results difficult (Eurostat, 2011).

For example, as far as sample design is concerned, while the common guidelines define a national representative probability sample of the population living in private households in the country, the sample design is carried out differently in each country. In some countries the sample consists of a simple random selection of households, individuals or dwellings, while in others a more complex multi-stage process is applied.

EU-SILC longitudinal component the persons who were selected initially are interviewed for a period of four years, equal to the duration of each panel (Eurostat, 2013).

3.5 Small Area Estimation of Poverty Indicators Using the Fay-Herriot model

Indicators of poverty have an important territorial dimension associated with the need to take into account regional and local differences in the construction of the system of these indicators (SAMPLE, 2009). Many countries around the world decentralize decision-making, resources and responsibilities to lower levels of government, and as a result, regional and local governments have more opportunities to tackle poverty. Thus, in order to ensure a good allocation of public resources, a system of indicators of poverty and social exclusion at regional and local level is necessary.

Most of the surveys of living conditions (sample surveys, Census and registers) provide high quality information only at the country or regional level since they are not large enough to support accurate estimates for small areas. The small sample size at the lower level of spatial aggregation leads to a large variation in the estimates obtained and therefore lower reliability. In fact, most current Eurostat and national surveys are planned to assure accuracy at NUTS 2 or 3 levels (e.g., Labor Force Survey, Household Budget Survey, Multipurpose Survey) (Pratesi, 2016). In Greece, for example, the minimum effective sample size of EU-SILC, which is 8250 households (4750 cross-sectional and 3500 Longitudinal) according to Eurostat (2018), aims to provide accurate estimates at regional level (NUTS 2). Therefore, small area estimation techniques for measuring poverty at local level are required that "borrow strength" across areas through linking models and auxiliary information such as Censuses and administrative data (Molina and Rao, 2010). Through using this information from other sources, it is possible to estimate distribution parameters with smaller variance than in the case of direct estimation.

FGT poverty measures for small areas. Suppose a population of size N is divided into D small areas of size $N_1, N_2, \dots, N_i, \dots, N_D$. Let E_{ij} be the income/expenditure for individual j in small area i ($j = 1,2,\dots,N_i$ and $i = 1,2,\dots,D$). Then the FGT poverty measures P_a given by formula (3.2.5), are written for each small area i as (Molina and Rao, 2010):

$$P_{ai} = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{z-E_{ij}}{z}\right)^a I\left(E_{ij} < z\right) \tag{3.5.1}$$

with $$I\left(E_{ij} < z\right) = \begin{cases} 1, & \text{if } E_{ij} < z \text{ (population unit under poverty)} \\ 0, & \text{otherwise} \end{cases} \tag{3.5.2}$$

For $a = 0,1,2$ the formula (3.5.1) gives the headcount ratio,

$$P_{0i} = \frac{1}{N_i}\sum_{j=1}^{N_i} I\left(E_{ij} < z\right) \tag{3.5.3}$$

the poverty gap index,

$$P_{1i} = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{z-E_{ij}}{z}\right) I\left(E_{ij} < z\right) \tag{3.5.4}$$

and the squared poverty gap index,

$$P_{2i} = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{z-E_{ij}}{z}\right)^2 I\left(E_{ij} < z\right) \tag{3.5.5}$$

respectively, for the small area i.

Direct estimator of FGT poverty measures for small areas. Suppose a random sample of size $n < N$ is drawn from the population according to a specified sampling design and $n_1, n_2, \ldots, n_D$ is the sample size of the selected units in each small area $i = 1,2,\ldots,D$. Let s_i be the set of units selected in the sample for the small area i and w_{ij} be the sampling weight of individual j from area i[33]. The basic direct estimators of the FGT measures are defined as (Molina and Rao, 2010):

$$\hat{P}_{ai}^{DIR} = \frac{1}{\hat{N}_i}\sum_{j\in s_i} w_{ij}\left(\frac{z-E_{ij}}{z}\right)^a I\left(E_{ij} < z\right), a = 0,1,2\ , i = 1,2,\ldots,D \tag{3.5.6}$$

where

$$\hat{N}_i = \sum_{j\in s_i} w_{ij} \tag{3.5.7}$$

is the direct estimator of the population size N_i of the i-th small area.

If the sampling weights w_{ij} do not depend on the unit j[34] then formula (3.5.6) reduces to the unweighted mean that is given by:

$$p_{ai} = \frac{1}{n_i}\sum_{j\in s_i}\left(\frac{z-E_{ij}}{z}\right)^a I\left(E_{ij} < z\right), a = 0,1,2\ , i = 1,2,\ldots,D \tag{3.5.8}$$

Indirect estimators of FGT poverty measures based on a Fay-Herriot model. Direct estimators of poverty measures in small areas are not accurate enough due to the limited sample sizes in these areas. As already detailed in chapter 2, to obtain reliable estimators for small domains or geographical areas it is necessary to appeal to small area techniques. Among many available small area methods, the Fay-Herriot model (2.3.8) is used widely in practice to estimate poverty rates in small areas.

According to formula (2.3.6), the Fay-Herriot area level model links the parameters of interest P_{ai} for all the areas $i = 1,2,\ldots,D$ through a linear model as:

$$P_{ai} = \boldsymbol{x}_i^T\boldsymbol{\beta} + u_i, i = 1,2,\ldots,D \tag{3.5.9}$$

---

[33] It holds that $w_{ij} = \frac{1}{\pi_{ij}}$, where π_{ij} is the inclusion probability for individual j in area i.

[34] As for example happens under simple random sampling within areas where $w_{ij} = \frac{n_i}{N_i}$.

where $\boldsymbol{x_i} = (x_{1i}, x_{2i}, \dots, x_{pi})$ is a vector of covariates (area-specific auxiliary data) for domain i, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ is the $p \times 1$ vector of regression coefficients and u_i's are area specific random effects. Fay-Herriot model assumes that $\hat{P}_{ai}^{DIR}$ is design-unbiased, with:

$$\hat{P}_{ai}^{DIR} = P_{ai} + e_i,\ i = 1,2,\dots,D \tag{3.5.10}$$

where e_i is the sampling error associated with the direct estimates of each small area i. Combining (3.5.9) and (3.5.10) we obtain the linear mixed model:

$$\hat{P}_{ai}^{DIR} = \boldsymbol{x}_i{}^T\boldsymbol{\beta} + u_i + e_i, i = 1,2,\dots,D \tag{3.5.11}$$

We assume that $u_i \overset{iid}{\sim} (0, \sigma_u^2)$ and $e_i \overset{ind}{\sim} (0, \psi_i)$, where the sampling variances ψ_i, $i = 1,2,\dots,D$, are supposed to be known.

The best linear unbiased predictor (BLUP) of $P_{ai} = \boldsymbol{x}_i{}^T\boldsymbol{\beta} + u_i$ under model (3.5.11) can be expressed as a weighted combination of the direct $\hat{P}_{ai}^{DIR}$ and the regression-synthetic estimators $\boldsymbol{x}_i{}^T\tilde{\boldsymbol{\beta}}$, that is (Guadarrama, Molina, Rao, 2014):

$$\tilde{P}_{ai}^{F-H} = \gamma_i\hat{P}_{ai}^{DIR} + (1-\gamma_i)\boldsymbol{x}_i{}^T\tilde{\boldsymbol{\beta}}, i = 1,2,\dots,D \tag{3.5.12}$$

with weight

$$\gamma_i = \sigma_u{}^2/(\psi_i + \sigma_u{}^2) \tag{3.5.13}$$

and

$$\tilde{\boldsymbol{\beta}} = (\textstyle\sum_{i=1}^{D}\gamma_i\boldsymbol{x}_i\boldsymbol{x}_i{}^T)^{-1}\sum_{i=1}^{D}\gamma_i\boldsymbol{x}_i\,\hat{P}_{ai}^{DIR} \tag{3.5.14}$$

is the weighted least squares estimator of $\boldsymbol{\beta}$.

In practice, the variance σ_u^2 of the area effects u_i is unknown and needs to be estimated. As already mentioned in paragraph 2.4.1, common estimation methods are maximum likelihood (ML) and restricted maximum likelihood (REML). Therefore, replacing $\hat{\sigma}_u^2$ for $\sigma_u{}^2$ in (3.5.13) we obtain the empirical best linear unbiased predictor (EBLUP) of P_{ai}, denoted here as $\hat{P}_{ai}^{F-H}$ and is given by:

$$\hat{P}_{ai}^{F-H} = \hat{\gamma}_i\hat{P}_{ai}^{DIR} + \left(1-\hat{\gamma}_i\right)\boldsymbol{x}_i{}^T\hat{\boldsymbol{\beta}},\ i = 1,2,\dots,D \tag{3.5.15}$$

where $\hat{\gamma}_i$ and $\hat{\boldsymbol{\beta}}$ are the values of γ_i and $\tilde{\boldsymbol{\beta}}$ when $\sigma_u{}^2$is replaced by an estimator $\hat{\sigma}_u^2$.

Finally, for unsampled domains ($n_i = 0$) the poverty rate is estimated using only auxiliary variables without sample data:

$$\hat{P}_{ai}^{F-H} = \boldsymbol{x}_i{}^T\tilde{\boldsymbol{\beta}},\ i = 1,2,\dots,D \tag{3.5.16}$$

Estimates obtained in this way are called synthetic (Rao and Molina, 2015).

The Fay-Herriot model offers many advantages for estimating poverty. A key advantage is that it requires only area level auxiliary information and therefore avoids the confidentiality issues associated with micro-data. Also, due to the aggregation of data, it is not much affected by isolated unit level outliers. Furthermore, as mentioned by Guadarrama et al. (2014), when estimator $\hat{P}_{ai}^{DIR}$ is inefficient for small area i (that is, with a large sampling variance ψ_i compared to the unexplained between-area variability σ_u^2), γ_i becomes small and $\tilde{P}_{ai}^{F-H}$ borrows more strength from the other areas through the regression-synthetic estimator $\boldsymbol{x}_i{}^T\tilde{\boldsymbol{\beta}}$. On the other hand, when estimator $\hat{P}_{ai}^{DIR}$ is efficient, γ_i is large and $\tilde{P}_{ai}^{F-H}$ attaches more weight to the direct estimator. Thus, a Fay-Herriot estimator automatically borrows strength for the areas where it is needed.

However, according to Guadarrama et al., (2014) the model has some drawbacks. In order to estimate different indicators depending on a common continuous variable, separate modeling and search for good covariates for each index is required. Also, once the model is fitted at the area level, small area estimates $\hat{P}_{ai}^{F-H}$ cannot be further disaggregated for subdomains or subareas within the areas unless a new good model is found at that sub-area level. Furthermore, the estimation of MSE requires the normality of u_i and e_i that might not hold for very complex poverty indicators.

The Fay-Herriot model is used widely in practice to estimate poverty rates in small areas. For instance, the S.A.I.P.E project[35] has been implementing the Fay-Herriot model since 1993 to produce on an annual basis, model-based county estimates of poor school-age children in the United Sates. The U.S. Department of Education uses these estimates to allocate general funds to counties annually and then states distribute these funds among school districts. If θ_i is the parameter to be estimated for area i, then in this application $\theta_i = log(Y_i)$, where Y_i is the true poverty count of the i-th small area (county). Direct estimators $\hat{Y}_i$ were calculated as a 3-year weighted average of poor school-age children obtained from the March Supplement of the Current Population Survey (CPS) and area level auxiliary variables $\boldsymbol{x_i}$ were obtained from administrative records (Bell, 2009). Is worth noting that in the past, this procedure was done using data

[35] https://www.Census.gov/programs-surveys/saipe.htm

from the last Census, but poverty measurements have changed significantly over time, so the estimates were not reliable.

In addition, Molina and Morales (2009) applied the FH model to the 2006 Spanish SILC estimate poverty rates for the 52 Spanish provinces by gender. In the application, the estimated sampling variances of the direct estimators have been treated as the true variances. The auxiliary variables used were the domain proportions of individuals with Spanish nationality, in different age groups and in several employment categories. According to the results, based on the FH model there was a gain in efficiency of the EBLUP (the CVs of EBLUP estimators were smaller than those of direct estimators, especially for smaller domain) compared to the direct estimator for most domains.

Furthermore, Szymkowiak et al. (2017) in order to obtain efficient estimates of the poverty indicator at the level of subregions (NUTS 3) in Poland applied the EBLUP estimator based on the Fay–Herriot model. The auxiliary variables used were the percentage of single people aged over 25, the number of rooms per one household member, the percentage of households with a bathroom or shower, the percentage of households with two persons aged over 25 with no more than vocational education, population density and the ratio of people deregistered to the number of people registered for permanent residence in the subregion. According to the results of the study, the model used gave high–quality statistical outputs (very accurate estimates of poverty) which can be treated as indicators. That is, they can be used as a reliable criterion for assessing comparability of the poverty indicator over time and across areas.

4. Application to Poverty: Estimating Poverty in Greece Using Small Area Estimation Methods

4.1 Introduction

In Greece, systematic empirical research on poverty is relatively limited and rather recent. The main limiting factor has been the lack of appropriate statistical data, as well as conceptual and analytical problems encountered in such efforts (Mitrakos, 2016). The main sources of data for the collection and analysis of poverty in the case of Greece are the Household Budget Survey (HBS)[36] and the EU Survey of Income and Living Conditions (EU-SILC). The methodology of both surveys allows publishing results at a national level or, at most, for large geographical divisions (NUTS 2). Information for more detailed sections is not available because of the too small sample size, which leads to large mean square errors (MSE) of the obtained estimates. This work aims to estimate two of the FGT poverty measures (headcount ratio and poverty gap index) in Greece at a lower level of spatial aggregation than the one used so far, that is at the level of sub regions-NUTS 3 (*Nomoi*), using the small area estimation methodology.

In line with the framework developed in the previous chapters, SAE methods were used to estimate poverty in Greece at NUTS 3 level at two different times, in 2009 (shortly before the start of the Greek financial crisis[37]) and 2013 (during the crisis). Specifically, combining data from the EU-SILC survey and the national Greek Census, the Fay-Herriot model was applied to produce estimates for the percentage of the Greek population below the poverty line (headcount ratio) and for the poverty gap. The underlying small area estimation model used unit level survey responses (EU-SILC survey) but area level auxiliary variables (Census data) due to the restrictions on linking unit level survey and Census data.

---

[36] The Household Budget Survey (HBS) is a national survey collecting information from a representative sample of households, on households' composition, members' employment status, living conditions and, mainly, focusing on their members' expenditure on goods and services as well as on their income. The HBS was conducted for the first time during the years 1957 – 1958 and at five-year intervals thereafter, while from 2008 it has been conducted on an annual basis (Hellenic Statistical Authority, 2015).

[37] A full-blown financial crisis took place in Greece by the end of 2009 (Gibson, Palivos and Tavlas , 2014)

4.2 Research characteristics

4.2.1 Poverty measures selected.

4.2.1 Poverty measures selected. According to the framework developed in Chapter 3, in order to select the most appropriate poverty measure at least three issues need to be addressed. First, the appropriate metric of individual well-being must be determined. Second, a poverty line must be set below which individuals are considered poor and third, it is necessary to select an aggregation procedure to obtain a poverty rate for society as a whole. In the present study FGT poverty measures have been used[38]. For these measures the equivalized disposable household income[39] was defined as a metric of well-being and 60% of median household income was selected as a poverty line. As an aggregation procedure, the headcount ratio P_{0i} and the poverty gap index P_{1i} (given by 3.5.3 and 3.5.4 respectively) were selected.

Table 4.2.1 Key measures of poverty selected in the study

Measure	Metric of well-being	Poverty line	Aggregation
FGT	Equivalized disposable household income	60% of median household income	Headcount ratio P_{0i} and Poverty gap index P_{1i}

4.2.2 Estimation process. The target parameters to be estimated were the headcount ratio and the poverty gap index for the years 2009 and 2013 in Greece. The Horvitz-Thompson direct estimates $\hat{P}_{ai}^{DIR}$, $a = 0,1$ (given by 3.5.6) of the FGT measures P_{ai} , $a = 0,1$ was derived for each small area i (*Nomarhies*, $i = 1,2, \ldots ,54$) for the years 2009 and 2013 using the unit level data from the EU-SILC surveys of 2009 and 2013 respectively. The indirect estimates of these two FGT measures were derived for the years 2009 and 2013 based on a Fay-Herriot model (as analyzed in paragraph 3.5). Area-specific auxiliary data from the Greek Census of 2001 and 2011 were used to implement the Fay-Herriot model for the years 2009 and 2013 respectively. The variance σ_u^2 of the area specific random effects was estimated using restricted maximum likelihood (REML) and then the empirical best linear predictor (EBLUP) $\hat{P}_{ai}^{F-H}$ (given by 3.5.15) of P_{ai} , $a = 0,1$ was obtained.

---

[38] Details given in paragraphs 3.2 and 3.5.

[39] The definition is given in paragraph 3.3.

All computations were performed using the software R. Functions from the package *sae* (Molina and Marhuenda, 2015) as well as R functions which were developed during the ESSnet project in SAE (ESSnet, 2012b) were used. Specifically, from the package *sae* functions *eblupFH*() and *mseFH*() were used and the functions *mixed.area.sae*() and *diagnostic*() from the ESSnet project.

4.2.3 NUTS in Greece. The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU and the UK for the purpose of the collection, development and harmonization of European regional statistics, for socio-economic analyses of the regions and for framing of EU regional policies. The NUTS classification subdivides the economic territory of the EU Member States into territorial units (regions) and includes three hierarchical levels. Each Member State is divided into NUTS 1 regions, which in turn are subdivided into NUTS 2 regions and then divided further into NUTS 3 regions. Each of these regions is allocated a specific code and name. The NUTS are complemented at the lower level by local administrative units. Districts and municipalities constitute a more detailed level than NUTS 3. These are called 'Local Administrative Units' (LAUs) and are the lower-level building blocks of the NUTS regions (Eurostat, 2018).

The NUTS classification can normally only be amended after at least three years have passed since the previous version. Amendments were established in 2006, 2010, 2013 and 2016. While only two countries had changes at the NUTS 1 level, six countries had changes at NUTS 2 and nine countries at the NUTS 3 level.

The current NUTS nomenclature, applicable from 1 January 2018, subdivides the economic territory of the European Union into 104 regions at NUTS 1 level, 281 regions at NUTS 2 level and 1,348 regions at NUTS 3 level. Below those, the Local Administrative Units (LAUs) have been defined. This LAU level consists of around 100,000 municipalities or equivalent units in the 28 EU Member States (as of 31 December 2017).

The correspondence between the NUTS levels and the national administrative units in Greece[40] based on NUTS 2016 and LAU 2017 are presented in table 4.2.2 below.

[40] From January 1, 2011, according to the Kallikratis program, the administrative system of Greece was drastically revised. The former system of 13 Regions, 54 Prefectures and 1033 Municipalities and communities was replaced by 7 decentralized administrations (*apokentromeni dioikisi)*, 13 regions (*perifereies*) and 325 municipalities (*dimoi*). The former Prefectures (*nomarhies*) largely still exist but

Table 4.2.2 Correspondence between the NUTS levels and the national administrative units in Greece based on NUTS 2016 and LAU 2017

NUTS level	**Units**	**Name**
NUTS 1	4	Great Geographical Areas (Megales Geografikes Perioches)
NUTS 2	13	Regions (Periferies)
NUTS 3	52	Groups of Regional Units (Omades Periferiakon Enotiton)
LAU	6,133	Municipal / Local Communes (Dimotikes / Topikes Koinotites)

Note: Adapted from Table 1 in *Regions in the European UNION*, (Eurostat, 2018, p.10).

The present survey was carried out using auxiliary information from the 2001 and 2011 national Greek Censuses. Between these two Censuses the administrative system of Greece changed. Therefore, some changes were made at the level of NUTS 3. The 2001 Census was carried out according to the Kapodistrias program, while the 2011 Census was carried out according to the Kallikratis program. In the present study, the administrative division used in the 2001 Census was adopted (Table 4.2.3)[41]. For the homogeneity and comparability of the survey results, some correspondences were made between the Prefectures (*Nomoi*) of 2001 and the Regional Units (*Perifereiakes Enotites*) of 2011 (Table 4.2.4).

are now called Regional Units (*Perifereiakes Enotites*) and form administrative and territorial constituent parts of the Regions (N. 3852/2010).

[41] In the present study, the Prefectures of Chalkidiki and Agion Oros were combined into one Prefecture. The Community of Mount Athos Monasteries (Agion Oros) is an autonomous territory with special status (*Ieri Koinotita*) under the Constitution and special arrangements as to the application of EU law to this territory exist.

Table 4.2.3 The Prefectures (Nomoi) of Greece (NUTS 3 Level) and their respective codes

Code	**Prefectures (*Nomoi*)**	**Code**	**Prefectures (*Nomoi*)**	**Code**	**Prefectures (*Nomoi*)**
300001	Etolia and Akarnania	300032	Thesprotia	300063	Florina
300003	Viotia	300033	Loannina	300064	Chalkidiki and Aghion Oros
300004	Evia	300034	Preveza	300071	Evros
300005	Evrytania	300041	Karditsa	300072	Xanthi
300006	Fthiotida	300042	Larissa	300073	Rodopi
300007	Fokida	300043	Magnissia	300081	Dodekanissos
300011	Argolida	300044	Trikala	300082	Kyklades
300012	Arkadia	300051	Grevena	300083	Lesvos
300013	Achaia	300052	Drama	300084	Samos
300014	Ilia	300053	Imathia	300085	Chios
300015	Korinthia	300054	Thessaloniki	300091	Iraklio
300016	Lakonia	300055	Kavala	300092	Lassithi
300017	Messinia	300056	Kastoria	300093	Rethymno
300021	Zakynthos	300057	Kilkis	300094	Chania
300022	Kerkyra	300058	Kozani	300101	Prefecture of Athens
300023	Kefallinia	300059	Pella	300102	Prefecture of East Attiki
300024	Lefkada	300061	Pieria	300103	Prefecture of West Attiki
300031	Arta	300062	Serres	300104	Prefecture of Pireas

Note: The administrative division used in the 2001 national Greek Census was adopted.

Table 4.2.4 Correspondences between Prefectures (*Nomoi*) of 2001 national Greek Census and Regional Units (*Perifereiakes Enotites*) of 2011 national Greek Census

<table>
<tr><th>Prefectures (Nomoi) of 2001 Census</th><th>Regional Units (Perifereiakes Enotites) of 2011 Census</th></tr>
<tr><td>Prefecture of Athens</td><td>Kentrikos Tomeas Athinon
Voreios Tomeas Athinon
Dytikos Tomeas Athinon
Notios Tomeas Athinon</td></tr>
<tr><td>Prefecture of Pireas</td><td>Perifereiaki Enotita Pireos
Perifereiaki Enotita Nison</td></tr>
<tr><td>Kavala</td><td>Perifereiaki Enotita Kavalas
Perifereiaki Enotita Thasou</td></tr>
<tr><td>Magnisia</td><td>Perifereiaki Enotita Magnisias
Perifereiaki Enotita Sporadon</td></tr>
<tr><td>Kefallinia</td><td>Perifereiaki Enotita Kefallinia
Perifereiaki Enotita Ithakis</td></tr>
<tr><td>Samos</td><td>Perifereiaki Enotita Samou
Perifereiaki Enotita Ikarias</td></tr>
<tr><td>Lesvos</td><td>Perifereiaki Enotita Lesvou
Perifereiaki Enotita Limnou</td></tr>
<tr><td>Kyklades</td><td>Perifereiaki Enotita Androu
Perifereiaki Enotita Milou
Perifereiaki Enotita Thiras
Perifereiaki Enotita Keas-Kithnou
Perifereiaki Enotita Mykonou
Perifereiaki Enotita Naxou
Perifereiaki Enotita Syrou
Perifereiaki Enotita Tinou
Perifereiaki Enotita Parou</td></tr>
<tr><td>Dodekanissos</td><td>Perifereiaki Enotita Kalymnou
Perifereiaki Enotita Karpathou-Kasou
Perifereiaki Enotita Ko
Perifereiaki Enotita Rodou</td></tr>
<tr><td>Chalkidiki</td><td rowspan="2">Chalkidiki kai Agion Oros</td></tr>
<tr><td>Agion Oros</td></tr>
</table>

Note: For the remaining 44 prefectures not presented in the table there was no change between the 2001 and 2011 Censuses.

4.3 Data Sources

The model developed in this research was based on data from the EU-SILC survey and the national Greek Census. In particular, sample data for direct poverty estimates at NUTS 3 level for the years 2009 and 2013 come from the EU –SILC 2009 and 2013 surveys respectively (details are given in paragraph 4.3.1)[42]. The data of the auxiliary variables of the year 2009 and 2013 were derived from the national Greek Census of 2001 and 2011, respectively (details are given in paragraph 4.3.2).

4.3.1 EU-SILC Greece 2009 and 2013. The Greek Survey on Income and Living Conditions is part of the European Statistical Program and has replaced since 2003 the European Community Household Survey (ECHP). The survey is conducted by the Hellenic Statistical Authority (ELSTAT). The main objective of the survey (as already discussed in paragraph 3.4) is to study, both at European and national level, the living conditions of households in relation to their income. With the collected information, Greece examines specific socio-economic variables that affect the living conditions of the population. It calculates structural indicators for social cohesion and systematically produces statistics on income inequalities, inequalities in household living conditions, poverty, and social exclusion at national level (Hellenic Statistical Authority, 2012). The survey consists of two components the cross-sectional and the longitudinal. The first refers to a specific time period, and the second to the changes occurring in three- or four-years' time. The collection of the necessary data is achieved through questionnaires answered by a representative sample.

EU-SILC survey in Greece is based on a two-stage stratified sampling of households from a sampling frame, which has been created on the basis of the results of the 2001 Population Census and covers completely the reference population. The first level of area stratification in the sampling design is the geographical stratification based on the distribution of the total country area into thirteen standard administrative regions corresponding to the European NUTS 2 level. The two major city agglomerations of Greater Athens and Greater Thessalonica constitute separate major geographical strata. The second level of area stratification implies grouping of municipalities and communes in each NUTS 2 administrative region by degree of

[42] The EU-SILC microdata, for both the year 2009 and 2013, were made available for this study upon request to ELSTAT and with the assent of the Committee on Statistical Confidentiality.

urbanization, i.e., depending on the size of their population[43]. In both the year 2009 and 2013 the number of the final strata in the thirteen (13) geographical regions was 50. The Greater Athens Area was divided into 31 strata of about equal size (equal numbers of households) and the Greater Thessaloniki Area was divided into 9 equally sized strata. Thus, the total number of strata of the survey was 90 (Hellenic Statistical Authority, 2012).

Sampling units were selected in two stages. The primary units were the areas (one or more unified building blocks) and the ultimate sampling units selected in each sampling area were the households[44]. The final sample size for the year 2009 was 7,036 households and 18,035 members of those households (15,045 aged 16+). The average was calculated as 2.6 members per household (Hellenic Statistical Authority, 2009). In 2013, the survey was conducted on a final sample of 7,349 households and on 18,030 members of those households (15,318 aged 16+). The average was calculated as 2.5 members per household (Hellenic Statistical Authority, 2013).

The poverty line was calculated with its relative concept (poor in relation to others) and defined at 60% of the median total equivalized disposable income of the household, using modified OECD equivalized scale. The OECD modified scale gives a weight of 1.0 to the first adult., 0.5 to other persons aged 14 or over who are living in the household and 0.3 to each child aged under 14.

The total equivalized disposable income of the household is calculated as the total disposable income of the household divided by the equivalized household size (OECD equivalized scale). For the calculation of the total equivalized disposable income of the household, the total net income is taken into account (that is income after deducting taxes and social contributions) received by all members of the household. In particular the income components included in the survey were:

- Income from work
- Income from property
- Social transfers and pensions
- Monetary transfers from other households and

[43] The scaling of urbanization was designed in four groups: i) ≥ 30.000 inhabitants ii) 5.000-29.999 inhabitants iii) 1.000-4.999 inhabitants iv)0-999 inhabitants (Hellenic Statistical Authority, 2012).

[44] The definition of household that Eurostat recommends is used. Household is defined as a person living alone or a group of people who live together in the same dwelling and share expenditures including the joint provision of the essentials of living.

• Imputed income from the use of company car

Also, the equivalent available individual income[45] was considered as the total available income of the household after its division by the equivalent size of household. The equivalent size of household is calculated according to the modified scale of OECD.

The poverty line for 2009 amounted to 6,897.30 euro per person annually and 19.7% of the total population was at risk of poverty[46] (Hellenic Statistical Authority, 2009a). For 2013, the poverty line amounted to 5,023 Euros per person annually and 23.1% of the total population was at risk of poverty (Hellenic Statistical Authority, 2013a). The reference income period was the previous calendar year, that is 2008 for the results of the 2009 survey and 2012 for the results of the 2013 survey.

From this microdata, unit level income data were used to derive two new survey variables. A binary variable that indicates if a person is above the poverty line or not (1 if a person is below the poverty line and 0 otherwise) and a variable that gives the proportionate poverty gap in an area (prefecture) as defined in the formula 3.2.3. The definition of income and the poverty line mentioned above were used to create the new variables.

4.3.2 2001 and 2011 national Greek Census. The main purpose of the Census is to provide a common list of characteristics, listed under common rules and procedures, to ensure the comparability of population and dwellings across the European Union. The Population-Housing Censuses 2001 and 2011 in Greece were conducted by ELSTAT (it has taken place in all EU Member States with harmonized definitions) in order to collect statistical data on the main characteristics of dwellings, the number and composition of households and nuclear families, as well as on the demographic, social, educational and economic characteristics of the resident population of the Country (Hellenic Statistical Authority, 2014). The survey unit was the residential dwelling and the individual. Four units of measure were used: the

[45] It is pointed out that in the distribution per person it is considered that each member of the household has the same income that corresponds to the equivalized disposable income. This means that each member of the household enjoys the same standard of living. Therefore, in the distribution per person, the income attributed to each individual does not represent a monetary gain, but an indicator of a standard of living (Hellenic Statistical Authority, 2013a).

[46] The at-risk-of poverty rate (after social transfers) calculated as the percentage of persons (over the total population) with an equivalized disposable income below the at-risk of-poverty line.

number and the percentage (%) of dwellings[47], of households[48], of nuclear families[49] and of individuals. The aggregate results were released and presented in statistical tables up to the level of the Regional Unit (NUTS 3) and of the Municipality (LAU 1). In Greece, the Census is carried out every ten years.

The data of the Census (for both the year 2001 and 2011) used in the present survey are given on the official website of ELSTAT[50] in statistical tables at regional level of NUTS 3. These data were used as area-specific auxiliary variables (presented in detail in § 4.4) in order to apply the Fay-Herriot model and to produce indirect estimates of the percentage of the Greek population below the poverty line.

4.4 Horvitz-Thompson (H-T) direct estimates of the FGT measures $\boldsymbol{P_{ai}}$, $\boldsymbol{a = 0, 1}$ for the years 2009 and 2013

Fay and Herriot model consists of a sampling model for the direct estimates and a linking model for the parameters of interest. The direct estimator of a small area uses only the sample data from the target small area. In the present study sample data derived from the EU-SILC survey in Greece for the years 2009 and 2013. The formulas used to calculate the direct estimates of the FGT measures P_{ai} , $a = 0{,}1$ for the years 2009 and 2013 are given in the tables 4.4.1 and 4.4.2. The results of the direct estimates as well as the corresponding variances and coefficients of variations (CV) for the years 2009 and 2013 are given in Tables A3, A4, A5 andA6 in the Appendix. It should be noted that for the year 2009 the prefecture of Kefalonia dropped out of the estimation process as the numerical value of the direct estimate of headcount ratio P_{0i} for the specific prefecture was zero.

---

[47] The dwellings are divided into Conventional and Non-conventional dwellings. As conventional is considered the dwelling which is a permanent and independent structure that consists of at least one regular room and it is intended to be used as a dwelling of a household for at least one year. As non-conventional is considered the dwelling which is a structure from shoddy materials not necessarily intended for dwelling, found occupied during Census period (huts, cabins, shacks, shanties, caravans, houseboats etc.) (Hellenic Statistical Authority, 2011).

[48] Household is defined as the total number of persons permanently residing in a dwelling, conventional or not, irrespective of whether they are relatives or not. (Hellenic Statistical Authority, 2011).

[49] Nuclear Family is defined as two or more persons who live in the same household and who are related as husband and wife, as cohabiting partners, or as parent and child. Thus, a nuclear family comprises a couple without children, or a couple with one or more children, or a lone parent with one or more children. (Hellenic Statistical Authority, 2011).

[50] https://www.statistics.gr/en/2011-Census-pop-hous and https://www.statistics.gr/el/statistics/-/publication/SAM04/2001.

Table 4.4.1 Summary formulas for the Horvitz-Thompson (H-T) direct estimators of the FGT measures $\boldsymbol{P_{ai}}$, $\boldsymbol{\alpha = 0, 1}$ for the year 2009

N_i : population size of the *i*-th prefecture n_i: sample size of the *i*-th prefecture s_i: set of units selected in the sample for the *i*-th prefecture w_{ij}: sampling weight for the j unit in the *i*-th prefecture $i = 1, \ldots, 53$
Parameters to be estimated
• The headcount ratio $P_{0i} = \frac{1}{N_i}\sum_{j=1}^{N_i} I(E_{ij} < 5{,}023)$ • The poverty gap index $P_{1i} = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{5{,}023-E_{ij}}{5{,}023}\right) I(E_{ij} < 5{,}023)$ with $I(E_{ij} < 5{,}023) = \begin{cases} 1, & \text{if } E_{ij} < 5{,}023 \\ 0, & \text{otherwise} \end{cases}$ and $i = 1, \ldots, 53$
H-T direct estimators
• $\hat{P}_{0i}^{DIR} = \frac{1}{\hat{N}_i}\sum_{j \in s_i} w_{ij}\, I(E_{ij} < 5{,}023)$ • $\hat{P}_{1i}^{DIR} = \frac{1}{\hat{N}_i}\sum_{j \in s_i} w_{ij}\left(\frac{5{,}023-E_{ij}}{5{,}023}\right) I(E_{ij} < 5{,}023)$ where $\hat{N}_i = \sum_{j \in s_i} w_{ij}$ is the direct estimator of the population size N_i of the *i*-th prefecture.

Table 4.4.2 Summary formulas for the Horvitz-Thompson (H-T) direct estimators of the FGT measures $\boldsymbol{P_{ai}}$, $\boldsymbol{\alpha = 0, 1}$ for the year 2013

N_i : population size of the *i*-th prefecture n_i: sample size of the *i*-th prefecture s_i: set of units selected in the sample for the *i*-th prefecture w_{ij}: sampling weight for the *j* unit in the *i*-th prefecture $i = 1, \ldots, 54$
Parameters to be estimated
• The headcount ratio $P_{0i} = \frac{1}{N_i}\sum_{j=1}^{N_i} I(E_{ij} < 6{,}897.3)$ • The poverty gap index $P_{1i} = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{6{,}897.3-E_{ij}}{6{,}897.3}\right) I(E_{ij} < 6{,}897.3)$ with $I(E_{ij} < 6{,}897.3) = \begin{cases} 1, & \text{if } E_{ij} < 6{,}897.3 \\ 0, & \text{otherwise} \end{cases}$ and $i = 1, \ldots, 54$
H-T direct estimators
• $\hat{P}_{0i}^{DIR} = \frac{1}{\hat{N}_i}\sum_{j \in s_i} w_{ij}\, I(E_{ij} < 6{,}897.3)$ • $\hat{P}_{1i}^{DIR} = \frac{1}{\hat{N}_i}\sum_{j \in s_i} w_{ij}\left(\frac{6{,}897.3-E_{ij}}{6{,}897.3}\right) I(E_{ij} < 6{,}897.3)$ where $\hat{N}_i = \sum_{j \in s_i} w_{ij}$ is the direct estimator of the population size N_i of the *i*-th prefecture.

4.5 Auxiliary variables

The selection of auxiliary data related to the variables of interest is vital when applying SAE methods. As already mentioned in paragraph 2.5 the success of any model-based method depends on how good the auxiliary data are as predictors of the study variables. In the present study, an initial set of covariates was derived from the 2001 and 2011 Census data in Greece. A total of 19 auxiliary variables from the 2001 national Census in Greece and 32 auxiliary variables from the 2011 Census were examined[51] (presented in detail in Tables 4.5.1 and 4.5.2). The initial set of auxiliary variables was selected based on the factors that seem to influence poverty the most, according to the literature (Mitrakos, 2014; European Commission, 2010; Hellenic Statistical Authority, 2009a, 2013a) as well as based on past researches to estimate poverty using SAE methods (S.A.I.P.E project; Tzavidis, Salvati, Pratesi and Chambers, 2008; Molina and Morales, 2010; Molina and Rao, 2009; Salvati, Giusti and Pratesi, 2014; Szymkowiak et al. 2017). Some of the key factors that are considered to make a person more "at risk" of poverty are: unemployment or having a poor quality (i.e., low paid or precarious) job, low levels of education and skills, the size and type of family, gender, disability or ill-health, being a member of minority ethnic groups and living in a remote or very disadvantaged community. Specifically, in Greece, groups at high risk of poverty according to the data from EU-SILC (Hellenic Statistical Authority, 2009a, 2013a) include principally the unemployed, single-parent households with at least one dependent child, households with one adult over 65 years of age, economically inactive persons excluding pensioners, households with 3 or more adults with dependent children, households living in rented accommodation and children 0-17 years of age. Based on the above theoretical framework and the availability of Census data, the main types of variable selected include variables related to:

- demographic characteristics (i.e., age, gender, marital status)
- socioeconomic status (i.e., employment status, occupation)
- educational attainment (i.e., low, medium, high level of education)
- amenities of dwellings (i.e., lack of kitchen, bath, indoor flushing toilet, electricity, heating, internet)

[51] Using data from two different Censuses it was not possible to use exactly the same variables to construct the 2009 and 2013 estimation models. Many variables were present at regional level NUTS 3 in the 2011 Census but not in the 2001 Census.

- Others (i.e., number of rooms, cars, children, members in the household, tenure status).

The initial sets of auxiliary variables derived from the data of the 2001 and 2011 Census in Greece are presented in detail in Tables 4.5.1 and 4.5.2 respectively.

Table 4.5.1 Initial set of auxiliary variables from 2001 national Greek Census

Variable name	Label	Definitions
Educational attainment[52]		
X1 (peduc1)	Low education	Percentage of people per prefecture who belong to one of the following categories: Illiterate, completed pre-primary education, left primary school but knows reading & writing, primary school certificate. (level 0-1 of ISCED-97)
X2 (peduc2)	Medium education	Percentage of people per prefecture who belong to one of the following categories: Lower secondary school certificate (*gymnasio*), Technical college certificate (*TES*), Technical school certificate (*TEL*), Secondary education certificate (*lykeio*), Post-secondary education degree (*IEK, Kolegia*) (level 2-4 of ISCED-97)
X3 (peduc3)	High education	Percentage of people per prefecture who belong to one of the following categories: Certificate of high technical schools, Degree of Technical Education colleges (*TEI, KATE, KATEE, Anoteris scholis kai eklisiastikis ekpedefsis*) Higher Education Degree, Master's, PhD. (level 5-6 of ISCED-97)
X4 (peduc4)	Women with low level of education	Percentage of women per prefecture with low level of education (as defined in X1).

[52] Educational attainment of a person is the highest level of an educational programme the person has successfully completed and the study field of this programme. The educational classification used in the 2001 Census in Greece is the International Standard Classification of Education, ISCED 1997, coded according to the seven categories ISCED-97 (UNESCO, 2006).
Level 0: Pre-primary education.
Level 1: Primary education.
Level 2: Lower secondary education.
Level 3: (Upper) secondary education.
Level 4: Post-secondary non-tertiary education.
Level 5: First stage of tertiary education (not leading directly to an advanced research qualification)
Level 6: Second stage of tertiary education (leading to an advanced research qualification)

Variable name	Label	Definitions
X5 (peduc5)	Women with high level of education	Percentage of women per prefecture with high level of education (as defined in X3).
AGE		
X6 (under 15)	People aged under 15 years old.	Percentage of people per prefecture under 15 years old.
X7 (29-49)	People aged 29-49 years old.	Percentage of people per prefecture aged 29 to 49 years old.
X8 (over 50)	People aged over 50 years old.	Percentage of people per prefecture aged over 50 years old.
X9 (over 65):	People aged over 65 years old.	Percentage of people per prefecture aged over 65 years old.
Gender		
X10 (pe women)	Women	Percentage of women per prefecture
X11 (pe women 15-74)	Women aged 15-74.	Percentage of women per prefecture aged 15-74
Employment status[53]		
X12 (pe unemployed)	Unemployed	Percentage of unemployed people per prefecture
X13 (pe inactive)	Inactive	Percentage of inactive people per prefecture
Amenities of dwellings		
X14 (pe no-kitchen)	Lack of kitchen	Percentage of dwellings per prefecture without kitchen
X15 (pe no-electricity)	Lack of electricity	Percentage of dwellings per prefecture without electricity
X16 (pe no-bath)	Lack of bath or shower	Percentage of dwellings per prefecture without bath or shower
X17 (pe no -toilet)	Lack of indoor flushing toilet	Percentage of dwellings per prefecture without an indoor flushing toilet

[53] According to the 2001 national Greek Census the employment status is defined as follows (Hellenic Statistical Authority, 2001):

Economically active population are:

a) **Employed:** Persons aged 10 years or older (for comparability with previous Censuses), who during the week preceding the Census, declared: (a) that they worked, even for just one hour, for pay or profit, in cash or in kind (b) they were not at work but had a job or business from which they were temporarily absent and
b) **Unemployed**: Persons aged 10 and over (for comparability with previous Censuses) who during the week preceding the Census, declared that they are looking for work and are acting in this direction. These are: a) people who have lost their jobs for any reason, and b) "young people", i.e., people who have been looking for work for first time

Inactive are those persons who neither classified as employed nor as unemployed (younger than 10 years old, pensioners, capital income recipients, etc.)

Variable name	Label	Definitions
Others		
X18 (pe over 3members)	Households with three or more members per room	Percentage of households per prefecture with three or more members per room
X19 (pe rented)	Rented dwellings	Percentage of rented dwellings per prefecture

Table 4.5.2 Initial set of auxiliary variables from 2011 national Greek Census

Variable name	Label	Definitions
Educational attainment[54]		
X1 (peduc1)	Low education	Percentage of people per prefecture who belong to one of the following categories: Illiterate, completed pre-primary education, left primary school but knows reading & writing, primary school certificate. (level 0-1 of ISCED-97)
X2 (peduc2)	Medium education	Percentage of people per prefecture who belong to one of the following categories: Lower secondary school certificate (*gymnasio*, *epagelmatikes sholes*), Secondary education certificate (*lykeio*), Post-secondary education degree (*IEK*, *Kolegia*) (level 2-4 of ISCED-97)
X3 (peduc3)	High education	Percentage of people per prefecture who belong to one of the following categories: Certificate of high technical schools, Degree of Technical Education colleges (A*TEI*, *ASPAITE*, *higher vocational and equivalent schools*), Higher Education Degree, Master's, PhD. (level 5-6 of ISCED-97)
X4 (peduc4)	Women with low level of education	Percentage of women per prefecture with low level of education (as defined in X1).
X5 (peduc5)	Women with high level of education	Percentage of women per prefecture with high level of education (as defined in X3).
AGE		

[54] The educational classification used in the 2011 national Greek Census is the International Standard Classification of Education, ISCED 1997, as described in Table 4.5.1.

Variable name	Label	Definitions
X6 (under 15)	People aged under 15 years old.	Percentage of people per prefecture under 15 years old.
X7 (29-49)	People aged 29-49 years old.	Percentage of people per prefecture aged 29 to 49 years old.
X8 (over 50)	People aged over 50 years old.	Percentage of people per prefecture aged over 50 years old.
X9 (over 65):	People aged over 65 years old.	Percentage of people per prefecture aged over 65 years old.
Gender		
X10 (pe women)	Women	Percentage of women per prefecture
X11 (pe active women)	Economically active women	Percentage of people per prefecture who are women and economically active.
Employment status[55]		
X12 (pe unemployed)	Unemployed	Percentage of unemployed people per prefecture
X13 (pe inactive)	Inactive	Percentage of inactive people per prefecture
X14 (pe work less than 20h)	Work less than 20 hours per week	Percentage (of employed) persons per prefecture with less than 20 hours of employment per week
X15 (pe active high educ)	Economically active with a high level of education.	Percentage of people per prefecture who are economically active and have a high level of education.
X16 (pe active low educ)	Economically active with a low level of education.	Percentage of people per prefecture who are economically active and have a low level of education.

[55] According to the 2011 national Greek Census the employment status is defined as (Hellenic Statistical Authority, 2011):

Economically active population are:

a) **Employed**: Persons aged 15 years or older, who during the week preceding the Census, declared: (i) that they worked, even for just one hour, for pay or profit, in cash or in kind (ii) they were not at work but had a job or business from which they were temporarily absent and
b) **Unemployed**: Persons aged 15 and over who during the week preceding the Census, declared: (i) that they were without work i.e., they were neither employed nor self-employed, or (ii) they were currently available for work , i.e., they were ready to start working as salaried employees or self-employed during the week preceding the Census and for two weeks after the Census and (iii) they were seeking for a job, i.e., they had taken all the necessary steps to search for a salaried job or self-employment, within 4 weeks before the end of the week preceding the Census.

Inactive are those persons who neither classified as employed nor as unemployed (younger than 15 years old, pensioners, capital income recipients, etc.)

Variable name	Label	Definitions
X17	Economically active aged 15-54	Percentage of people per prefecture who are economically active and are aged 15 to 54 years.
Amenities of dwellings/households		
X18 (pe no-bath)	Lack of bath or shower	Percentage of dwellings per prefecture without bath or shower
X19 (pe no -toilet)	Lack of indoor flushing toilet	Percentage of dwellings per prefecture without an indoor flushing toilet
X20 (pe no-heating)	Lack of heating	Percentage of dwellings per prefecture without heating
X21 (pe no-energy source)	Lack of energy source for cooking	Percentage of households per prefecture without energy source for cooking
X22 (pe no-internet)	Lack of internet access	Percentage of households per prefecture without internet access
Occupations		
X23 (pe occup1)	People working in agriculture, forestry and fisheries	Percentage of people per prefecture working in agriculture, forestry and fisheries
X24 (pe occup2)	People working in construction	Percentage of people per prefecture working in construction
Others		
X25 (pe rented)	Rented dwellings	Percentage of rented dwellings per prefecture
X26 (pe density)	Less than 15 m^2 per household member	Percentage of dwellings per prefecture with a density of less than 15 square meters per household member
X27 (pe children)	Three or more children	Percentage of nuclear families per prefecture with three or more children
X28 (pe members)	Five or more members	Percentage of households per prefecture with five or more members
X29 (pe cars)	Three or more cars	Percentage of households per prefecture with three or more cars available for use by the household

Variable name	Label	Definitions
X30 (pe no_employed member)	Without employed member	Percentage of nuclear families per prefecture without an employed member
X31 (pe single parent))	Single-parent families	Percentage of single-parent nuclear families per prefecture
X32 (pe no cars)	Without car	Percentage of households per prefecture without any car

4.6 Model selection

In order to build the optimal small area model for estimating headcount ratio and poverty gap index in Greece for the years 2009 and 2013, a three-phase variable selection process was performed. In the first phase a correlation matrix[56] among the auxiliary variables (the initial set of auxiliary variables as presented in tables 4.5.1 and 4.5.2) themselves was analyzed. The purpose of the screening based on the correlation was to avoid multicollinearity[57]. The pairs of auxiliary variables with high correlation were examined and the decision to drop covariates was based on whether they were correlated with other covariates, how many other covariates, and how high the correlation was.

In the second phase as there was a large number of possibilities, a step-forward procedure was used starting with a null model, i.e., just an intercept and then one-by-one covariates were added until there was no improvement in terms of a selection criterion. That is, of the models produced, the acceptable ones were those where:

i) all included covariates were significant (p-values of the significance of each coefficient less than 10%), and no others were significant enough to enter the model and

ii) the sign next to each variable (that is, the sign of estimated model coefficients - beta parameter) was justified according to knowledge of the analyzed phenomena from the literature. For example, it is expected that when the percentage of people with low level of education increases, so does the level of poverty, so the symbol next to the variable corresponding to low level of education should be positive.

The models that were created from the above two phases are presented in the Tables 4.6.1 and 4.6.2 below. For each model are given: the estimated model

[56] The correlation matrices for both the variables in Table 4.5.1 and the variables in Table 4.5.2 are presented in the Appendix in Tables A1 and A2, respectively.
[57] Multicollinearity is described in paragraph 2.5.

coefficients, the p-value of the significance of each coefficient, the estimated variance of the area random effects and the information criteria.

Table 4.6.1 Coefficients (estcoef), estimated variance of the area random effects (refvar) and information criteria (goodness) for the selected models of 2009 for estimating headcount ratio using the EBLUP F-H model.

model 1

X1: Low education
X13: Inactive people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.6445719	0.2358351	-2.733147	0.0062732340
X1	0.6669005	0.1950621	3.418914	0.0006287168
X13	0.9636103	0.4520773	2.131517	0.0330465991

refvar

[1] 0.006066004

goodness

loglike	AIC	BIC	KIC	cAIC
50.65762	-93.31524	-85.43407	-89.31524	-142.12

model 2

X3: High education
X9: People aged over 65 years old

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	0.1952902	0.1217082	1.604578	0.10858676
X3	-1.2025178	0.6059707	-1.984449	0.04720583
X9	0.8758674	0.4743669	1.846392	0.06483525

refvar

[1] 0.00761452

goodness

loglike	AIC	BIC	KIC	cAIC
46.30926	-84.61852	-76.73736	-80.61852	-140.2801

model 3

X3: High education
X8: People aged over 50 years old

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	0.04806526	0.1664702	0.2887319	0.77278655
X3	-1.15838450	0.5850591	-1.9799446	0.04770976
X8	0.85161868	0.3869478	2.2008621	0.02774579

refvar
[1] 0.007251289
goodness

loglike	AIC	BIC	KIC	cAIC
47.01734	-86.03467	-78.15351	-82.03467	-140.2035

model 4

X3: High education
X13: Inactive people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.2739584	0.2733394	-1.002265	0.31621575
X3	-1.2767440	0.5486200	-2.327192	0.01995505
X13	1.1625201	0.4668536	2.490117	0.01277009

refvar
[1] 0.006908976

goodness

loglike	AIC	BIC	KIC	cAIC
47.65193	-87.30386	-79.42270	-83.30386	-140.019

model 5

X4: Women with low level of education
X13: Inactive people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.5906737	0.2348915	-2.514666	0.0119145280
X4	1.3239407	0.3944013	3.356836	0.0007883982
X13	0.8300211	0.4678586	1.774085	0.0760490130

refvar
[1] 0.006085631

goodness

loglike	AIC	BIC	KIC	cAIC
50.45334	-92.90668	-85.02551	-88.90668	-141.7848

model 6

X5: Women with high level of education
X8: People aged over 50 years old

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	0.03281612	0.1656222	0.1981385	0.84293673
X5	-2.28375403	1.2202232	-1.8715871	0.06126375
X8	0.86211157	0.3895618	2.2130292	0.02689563

refvar
[1] 0.007337971

goodness

loglike	AIC	BIC	KIC	cAIC
46.80652	-85.61305	-77.73188	-81.61305	-140.1484

model 7

X5: Women with high level of education
X9: People aged over 65 years old

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	0.1814469	0.1210129	1.499401	0.13376953
X5	-2.3608117	1.2704177	-1.858296	0.06312705
X9	0.8839501	0.4800075	1.841534	0.06554338

refvar
[1] 0.007712896

goodness

loglike	AIC	BIC	KIC	cAIC
46.06568	-84.13136	-76.25019	-80.13136	-140.1975

model 8

X5: Women with high level of education
X13: Inactive people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.2679076	0.2847407	-0.9408827	0.34676499
X5	-2.3978276	1.1777830	-2.0358823	0.04176218
X13	1.1194690	0.4832564	2.3165117	0.02053035

refvar
[1] 0.007112634

goodness

loglike	AIC	BIC	KIC	cAIC
47.01440	-86.02879	-78.14763	-82.02879	-139.627

model 9

X7: People aged 29-49 years old
X12: Unemployed people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	1.031971	0.2412750	4.277160	1.892931e-05
X7	-3.296210	0.8781181	-3.753720	1.742291e-04
X12	1.021656	0.4825359	2.117265	3.423735e-02

refvar
[1] 0.00691899

goodness

loglike	AIC	BIC	KIC	cAIC
48.07403	-88.14805	-80.26689	-84.14805	-140.7879

model 10

X8: People aged over 50 years old
X12: Unemployed people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.3335504	0.1480827	-2.252460	0.0242932050
X8	1.2724643	0.3560449	3.573887	0.0003517204
X12	1.0531642	0.4841761	2.175168	0.0296175357

refvar
[1] 0.006910123

goodness

loglike	AIC	BIC	KIC	cAIC
47.41095	-86.82189	-78.94072	-82.82189	-139.3732

model 11

X9: People aged over 65 years old
X12: Unemployed people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.1270452	0.1064636	-1.193320	0.23274401
X9	1.3874725	0.4337399	3.198858	0.00137973
X12	0.9944022	0.4962414	2.003868	0.04508423

refvar
[1] 0.007447361

goodness

loglike	AIC	BIC	KIC	cAIC
46.34432	-84.68863	-76.80747	-80.68863	-139.526

Note: Models created using auxiliary variables of the 2001 national Greek Census and EUSILC data 2009.

Table 4.6.2 Coefficients (estcoef), p-values, estimated variance of the area random effects (refvar) and information criteria (goodness) for the selected models of 2013 for estimating headcount ratio using the EBLUP F-H model.

model 1

X3: High education
X7: People aged 29-49 years old.
X8: People aged over 50 years old.
X10: Women

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	0.3499669	0.7205121	0.4857196	0.62716597
X3	-0.8615211	0.3831531	-2.2485034	0.02454411
X10	2.0280904	1.1853309	1.7109910	0.08708278
X7	-1.8260570	0.9280171	-1.9676976	0.04910285
X8	-1.2425470	0.3784289	-3.2834358	0.00102550

refvar

[1] 0.002781498

goodness

loglike	AIC	BIC	KIC	cAIC
69.91789	-127.83579	-115.90188	-121.83579	-162.4242

model 2

X3: High education
X9: People aged over 65 years old
X10: Women

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.5553104	0.5873989	-0.945372	0.344468998
X3	-1.1877284	0.3731643	-3.182856	0.001458301
X10	2.2192797	1.2120531	1.831009	0.067099251
X9	-0.7642944	0.2822611	-2.707757	0.006773966

refvar

[1] 0.003069774

goodness

loglike	AIC	BIC	KIC	cAIC
67.81948	-125.63895	-115.69403	-120.63895	-161.975

model 3

X3: High education
X10: Women
X27: Three or more children

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-1.0476843	0.6109129	-1.714949	0.08635461
X3	-0.6057321	0.3412786	-1.774890	0.07591604
X10	2.5270643	1.2335710	2.048576	0.04050357
X27	1.3623465	0.5627459	2.420891	0.01548251

refvar
[1] 0.00308968

goodness

loglike	AIC	BIC	KIC	cAIC
67.09213	-124.18427	-114.23935	-119.18427	-160.6818

model 4

X3: High education
X8: People aged over 50 years old
X10: Women

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.5164184	0.5871468	-0.8795388	0.379109212
X3	-1.1533124	0.3664038	-3.1476539	0.001645864
X10	2.3225548	1.2130879	1.9145809	0.055545978
X8	-0.6465717	0.2359344	-2.7404729	0.006135084

refvar
[1] 0.003052109

goodness

loglike	AIC	BIC	KIC	cAIC
67.90073	-125.80146	-115.85654	-120.80146	-161.9873

model 5

X3: High education
X6: People aged under 15 years old
X10: Women
estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-1.0572113	0.5567937	-1.898749	5.759752e-02
X3	-0.8417193	0.3112685	-2.704158	6.847770e-03
X6	2.3182869	0.5934169	3.906675	9.357492e-05
X10	2.1278159	1.1273337	1.887477	5.909626e-02

refvar
[1] 0.002503951

goodness

loglike	AIC	BIC	KIC	cAIC
71.22858	-132.45715	-122.51223	-127.45715	-163.9244

model 6

X1: Low education
X6: People aged under 15 years old
X10: Women

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-1.2022778	0.5339767	-2.251555	2.435042e-02
X1	0.4615928	0.1368021	3.374164	7.404020e-04
X10	1.7174816	0.9855418	1.742677	8.138997e-02
X6	2.6378629	0.5847003	4.511479	6.437725e-06

refvar
[1] 0.002220133

goodness

loglike	AIC	BIC	KIC	cAIC
73.06588	-136.13177	-126.18685	-131.13177	-164.9144

model 7

X1: Low education
X10: Women
X27: Three or more children

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-1.0992219	0.6014648	-1.827575	0.067613382
X1	0.2976846	0.1477462	2.014838	0.043921656
X27	1.4815284	0.5475806	2.705590	0.006818321
X10	2.2053309	1.1304100	1.950912	0.051067474

refvar
[1] 0.002960068

goodness

loglike	AIC	BIC	KIC	cAIC
67.56141	-125.12283	-115.17791	-120.12283	-160.4991

model 8

X10: Women
X23: People working in agriculture, forestry and fisheries
X27: Three or more children

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.9753975	0.5786267	-1.685711	0.091851431
X23	0.2250432	0.1096575	2.052238	0.040146494
X27	1.4498822	0.5473711	2.648810	0.008077566
X10	2.1388301	1.1168884	1.914990	0.055493777

refvar
[1] 0.002939773

goodness

loglike	AIC	BIC	KIC	cAIC
67.63776	-125.27551	-115.33059	-120.27551	-160.4294

model 9

X12: Unemployed
X23: People working in agriculture, forestry and fisheries
X25: Rented dwellings

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.08043712	0.1282952	-0.6269691	0.530679466
X23	0.39625567	0.1494372	2.6516530	0.008009881
X25	0.54872977	0.2905097	1.8888519	0.058911678
X12	0.85177311	0.3928180	2.1683659	0.030130855

refvar
[1] 0.003359412

goodness

loglike	AIC	BIC	KIC	cAIC
66.16786	-122.33571	-112.39079	-117.33571	-160.8776

Note: Models created using auxiliary variables of the 2011 national Greek Census and EUSILC data 2013

In the third phase the final model for each year was chosen according to three information criteria. As already mentioned and analyzed in paragraph 2.6.1 several criteria can be applied to select the "best" model. In the present study three information criteria were used as comparison measures in order to select the final model for each year. These criteria were AIC (Akaike, 1974), BIC (Schwartz, 1978) and cAIC (Vaida and Blanchard, 2005)[58].The results of the information criteria for each model in Tables 4.6.1 and 4.6.2 are presented in Tables 4.6.3 and 4.6.4.

[58] Details of these criteria are given in paragraph 2.6.1.

Table 4.6.3 Information criteria AIC, BIC, cAIC for the selected models for estimating headcount ratio for the year 2009

Final Models	**Information Criteria**		
	AIC	BIC	cAIC
model 1	**-93.31524**	**-85.43407**	**-142.12**
model 2	-84.61852	-76.73736	-140.2801
model 3	-86.03467	-78.15351	-140.2035
model 4	-87.30386	-79.42270	-140.019
model 5	-92.90668	-85.02551	-141.7848
model 6	-85.61305	-77.73188	-140.1484
model 7	-84.13136	-76.25019	-140.1975
model 8	-86.02879	-78.14763	-139.627
model 9	-88.14805	-80.26689	-140.7879
model 10	-86.82189	-78.94072	-139.3732
model 11	-84.68863	-76.80747	-139.526

Note: The figures displayed in black denote the best values for AIC, BIC and cAIC. Also, models created using auxiliary variables of the 2001 national Greek Census and EUSILC data 2009.

Table 4.6.4 Information criteria AIC, BIC, cAIC for the selected models for estimating headcount ratio for the year 2013

Final Models	**Information Criteria**		
	AIC	BIC	cAIC
model 1	-127.83579	-115.90188	-162.4242
model 2	-125.63895	-115.69403	-161.975
model 3	-124.18427	-114.23935	-160.6818
model 4	-125.80146	-115.85654	-161.9873
model 5	-132.45715	-122.51223	-163.9244
model 6	**-136.13177**	**-126.18685**	**-164.9144**
model 7	-125.12283	-115.17791	-160.4991
model 8	-125.27551	-115.33059	-160.4294
model 9	-122.33571	-112.39079	-160.8776

Note: The figures displayed in black denote the best values for AIC, BIC and cAIC. Also, models created using the auxiliary variables of the 2011 national Greek Census and EUSILC data 2013.

According to the above tables the smallest value for all three information criteria corresponds to model 1 (Table 4.6.3) for estimating headcount ratio for the year 2009 and to model 6 (Table 4.6.4) for the year 2013. Therefore, these two models were selected as the final models for carrying out the estimation process. For reasons of comparability of the results these models were also used to estimate the poverty gap index.

4.7 Model Diagnostics and Evaluation

Small area estimation approaches depend to a large extent on the quality of the models used. Wrong specifications can lead to a strong bias of the estimators and accordingly to a misleading information base in the applications (Rao and Wu, 2001). Therefore, after applying small-area estimation, the model chosen must be carefully examined and checked for a violation of the underlying assumptions and a possible bias. As already discussed in section 2.6.2, different diagnostic tools can be applied to assess the fit and the performance of a model-based estimator as well as to check the reliability of the results.

The procedure followed in order to evaluate the final models was:

i. **Bias diagnostic**[59] was used to test whether a potential bias exists. In order to examine whether a substantial bias exists a scatter plot is produced, which illustrates the relation between direct and model-based estimates.

> On one hand the SAE estimates (X-axis) are plotted on a cartesian plane against the direct estimates (Y-axis) to verify if there is a departure of the regression line between model based and direct estimates from y = x. On the other hand, a parametric test for the slope and for the intercept is carried out to check the unbiasedness of the model predictions. It would be expected that the slope of the regression line is not significantly different from 1 and there is a very small intercept term, not significantly different from 0 (ESSnet, 2012b, p. 107).

ii. **Goodness of fit diagnostic**[60] was used to test whether the model estimates are close to the direct estimates. To evaluate this Brown et al. (2001) propose a Wald statistic[61], calculated as the sum of the squared differences between direct and model-based estimates (over all areas) weighted by their variances. Under the hypothesis that the model-based estimates are equal to the expected values of the direct estimates and provided the sample sizes in the small areas are sufficient to justify central limit assumptions, W will have a χ^2 distribution with

[59] This diagnostic tool was applied to the ESSnet project (2012d) in the case studies developed by Italy (p. 59), Poland (p. 128), Switzerland (p.167) and France (p. 28).

[60] This diagnostic tool was applied to the ESSnet project (2012d) in the case studies developed by Netherlands (p. 80) and France (p. 28) as well as to the ESSnet project (2012a) in the case study developed by the United Kingdom (p. 86).

[61] The Wald statistic is given by the equation: $W = \sum_i \frac{\left(\hat{\theta}_i^{DIR} - \hat{\theta}_i^{M-B}\right)^2}{\left(V\left(\hat{\theta}_i^{DIR}\right) + V\left(\hat{\theta}_i^{M-B}\right)\right)}$

degrees of freedom equal to the number of small areas in the population. As a check for the lack of bias of the model estimates, W is compared with the quantiles of χ^2 distribution.

iii. **Coverage diagnostic** was used in order to evaluate the validity of the confidence intervals generated by the model-based estimation procedure (Ambler et al., 2001). As noted by Brown et al. (2001, p. 7):

> 95% Confidence intervals for the direct estimates should contain the "truth" 95% of the time. So should the confidence intervals surrounding model-based estimates. Both sets of intervals are adjusted so that their chance of overlapping should be 95% and it is counted how often they actually do overlap. Assuming that the estimated coverage of the direct confidence intervals is correct, comparing the counts to the Binomial distribution provides a non-parametric significance test of the bias of model estimates relative to their precision.

iv. **Residual analysis**[62] was used to test whether the model assumptions are satisfied, i.e., the assumption of a normal distribution of the sampling errors defined in equation (3.5.10) and the area-specific random effects of (3.5.9) (ESSnet, 2012c, p.53; Tzavidis et al. 2018). To test the hypothesis of a normal distribution of the sampling errors, the tools used were:

- Normal quantile-quantile plot (Q-Q plot) for the standardized residuals[63]. This plot illustrates the relation between the sample quantiles of the standardised residuals and the theoretical quantiles of a normal distribution. If the standardised residuals are normally distributed, they will lie on a straight line (Eurostat, 2019, p. 21; Szymkowiak et al. 2017).
- Shapiro-Wilk test for normality. If the normal Q-Q plot for the standardized residuals of the model-based estimation indicates a slight deviation from the normal distribution, then a Shapiro-Wilk test for normality is a further tool to check whether the null hypothesis of normality can be rejected (Eurostat, 2019, p. 21).
- Plot of model-based estimates versus residuals. This plot helps to examine if the assumption that sampling errors have a constant variance is satisfied. A

[62] This diagnostic tool was applied to the ESSnet project (2012d) in the case studies developed by France (p. 25), Italy (p. 64), Netherlands (p. 81) and Switzerland (p. 163) as well as to the ESSnet project (2012a) in the case study developed by United Kingdom (p. 85).

[63] Standardized residuals r_i, are given from the formula $r_i = (\psi_i + \sigma_v^2)^{-\frac{1}{2}} (\hat{\theta}_i^{DIR} - \overline{\boldsymbol{X}}_i^{T} \boldsymbol{\beta})$ with $i = 1,2, \ldots D$.

systematic pattern in this plot indicates that the assumption is not satisfied (Tzavidis et al., 2018).

The same tools were used to analyze the distribution of random effects.

v. After the above diagnostic tests have been performed the quality of the model-based estimates has to be checked. One of the most relevant measures of the quality of an estimator is the Mean square error (MSE) (Tzavidis et al. 2018). The ultimate aim of course is to produce estimates for the small areas that have lower mean square errors (MSEs) than the direct estimates. Each small area estimate in the present study is accompanied by an estimate of its MSE. In addition, a gain-in-precision index (GIP1) is calculated from the equation (Szymkowiak et al., 2017):

$$GIP1_i = \frac{\sqrt{\psi_i}}{\sqrt{MSE(\hat{\theta}_i)}} \tag{4.7.1}$$

where ψ_i is the sampling variance of the *i*-th area and $MSE(\hat{\theta}_i)$ the estimated MSE of the EBLUP in the *i*-th area. Their square roots represent the standard errors. This ratio was calculated in order to analyze the efficiency gain achieved by the application of small area estimation methods. Specifically, it shows by what factor the standard error was reduced by the small area estimator compared to the direct estimator. In addition, GIP1 ratio was illustrated in relation to the area-specific sample size in order to detect whether an increasing number of observations per area results in a decreased improvement by using the small area estimation approach.

vi. Finally, another measure of the quality of an estimator is the coefficient of variation (CV). As with MSE, the ultimate aim is to produce estimates for small areas with lower CVs than direct estimates (Molina and Morales, 2009; Molina and Rao, 2010). Ideally, the majority of coefficients of variation should be below 20% for estimates to be precise and considered adequate for publishing (ONS, 2006; Molina and Marhuenda, 2015). In the present study the ratio (GIP2) of coefficients of variation (CV's) of direct estimators over Fay and Herriot estimators was calculated and illustrated in relation to the area-specific sample size. The formula for the GIP2 is given by (Molina and Rao, 2010):

$$GIP2_i = \frac{{CV_i}^{DIR}}{{CV_i}^{F-H}} \tag{4.7.2}$$

where CV_i^{DIR} is the coefficient of variation of the i-th area for the direct estimator and CV_i^{F-H} is the coefficient of variation of the i-th area for the Fay and Herriot estimator.

The numeric values of the EBLUP Fay and Herriot estimators as well as the corresponding mean square errors (MSE), the coefficients of variation (CV), the GIP1 and GIP2 ratios are given in Tables A7-A14 in the Appendix.

4.7.1 Model Diagnostics for the estimation of FGT measures in Greece for the year 2013. For the estimation of FGT measures P_{ai}, $a = 0{,}1$ model 6 (Table 4.6.4) was selected. In model 6 there were three covariates: Percentage of people per prefecture with low education (X1), percentage of people per prefecture aged under 15 years old (X6) and percentage of women per prefecture (X10). All β (beta) parameters (except intercept) have a positive sign (Table 4.7.1), which means that the increase of any one of those three indicators in an area is associated with an increase of the poverty rate in this unit. Indeed, according to the literature and the results of the EU-SILC survey, among the groups at high risk of poverty are people with low level of education, women and people under 15.

Table 4.7.1 Coefficients for the final selected model for estimating the headcount ratio and poverty gap index for the year 2013 using F-H model

Headcount ratio (P_{0i})				
	beta	std.error	tvalue	pvalue
(Intercept)	-1.2022778	0.5339767	-2.251555	2.435042e-02
X1 low education	0.4615928	0.1368021	3.374164	7.404020e-04
X10 women	1.7174816	0.9855418	1.742677	8.138997e-02
X6 people under 15	2.6378629	0.5847003	4.511479	6.437725e-06
Poverty gap index (P_{1i})				
	beta	std.error	tvalue	pvalue
(Intercept)	-1.1002477	0.25503162	-4.314162	1.602096e-05
X1 low education	0.2229218	0.06883168	3.238651	1.20096e-03
X10 women	1.7926809	0.47359988	3.785223	1.535711e-04
X6 people under 15	1.3100994	0.29093003	4.503142	6.695596e-06

i) In order to examine whether a substantial bias exists, the graphical diagnostic suggested by Brown et al. (2001) (presented in section 4.7 (i)) was produced and illustrated in Figures 4.7.1 and 4.7.2 for the estimate of headcount ratio and poverty gap, respectively. In these figures the bisector is shown in black, whereas the regression line is red. Also, a goodness of fit diagnostic proposed by Brown et al. (2001) (as described in section 4.7 (ii)) was produced[64], and the results are given in Tables 4.7.2 and 4.7.3.

In the case of the headcount ratio, as shown in the Figure 4.7.1 the intercept of the linear regression is $\beta_0 = -0.01696$ and the parameter for the slope is $\beta_1 = 1.14137$. The intercept is very close to zero and the slope estimates do not differ much from 1. It seems that the cloud of points spreads along the line Y=X, so there is a strong assumption of lack of bias. The goodness of fit diagnostic, as shown in Table 4.7.2, accept null hypothesis that the Fay-Herriot estimates are close to the direct estimates when the direct estimates are good.

In the case of the poverty gap, as shown in the Figure 4.7.2 the intercept of the linear regression is $\beta_0 = 0.00278$ and the parameter for the slope is $\beta_1 = 1.09326$. It seems that the cloud of points spreads along the line Y=X, so there is a strong assumption of lack of bias. The goodness of fit diagnostic, as shown in Table 4.7.3, accept null hypothesis that the Fay-Herriot estimates are close to the direct estimates.

Therefore, it appears that both the model for estimating headcount ratio and the model for estimating poverty gap are unbiased.

[64] For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used. This R function was developed for the application of model bias diagnostic proposed by Brown et al. (2001).

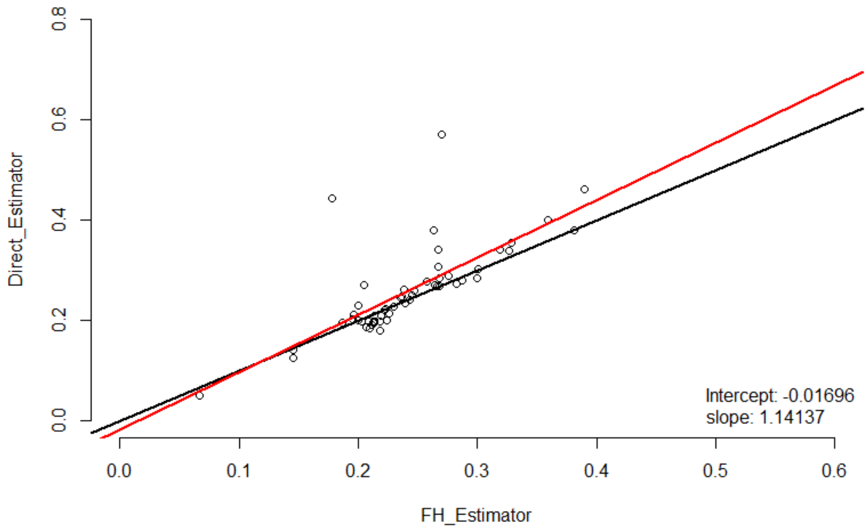

Figure 4.7.1 Relation between direct estimates and Fay-Herriot estimates of the headcount ratio in Greece for the year 2013

Table 4.7.2 The values of the empirical Wald test (W), of the theoretical χ^2 (c_alfa1), the p-value and the test result for the Fay and Herriot model estimator of the headcount ratio in Greece for the year 2013

method	W	c_alfa1	p-value	results
eblup.area	15.10262	70.99345	6.399869e-08	Accept H_0: E(Direct estimates) = Model based Estimates

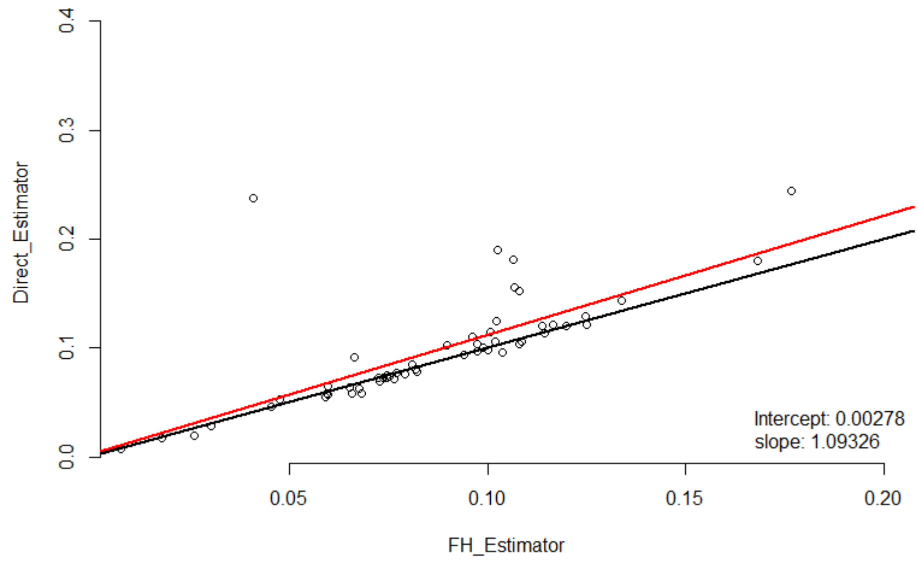

Figure 4.7.2 Relation between direct estimates and Fay-Herriot estimates of the poverty gap in Greece for the year 2013

Table 4.7.3 The values of the empirical Wald test (W), of the theoretical χ^2 (c_alfa1), the p-value and the test result for the Fay and Herriot model estimator of the poverty gap in Greece for the year 2013

method	W	c_alfa1	p-value	results
eblup.area	16.98440	70.99345	5.877562e-07	Accept H0: E(Direct estimates) = Model based Estimates

ii) In order to evaluate the validity of the confidence intervals generated by the Fay and Herriot model a coverage diagnostic (analyzed in section 4.7 (iii))[65] was used. The results are given in Tables 4.7.4 and 4.7.5 for the headcount ratio and the poverty gap, respectively. Also, the numerical values of the confidence intervals are given in Tables A7 and A8 in the Appendix, for the Fay and Herriot estimates of the headcount ratio and poverty gap respectively. An illustration of the results is presented in Figures 4.7.3 and 4.7.4. In both cases the null hypothesis that the overlap is 95% is accepted. This means that the confidence intervals generated by Fay and Herriot model are valid.

[65] For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used.

Table 4.7.4 The values of the empirical z, of the theoretical z (z_teo), the p-value, the overlapped areas, the overlap rate (f_sovrap), and the result of the test for the Fay and Herriot model estimator of the headcount ratio in Greece for the year 2013

method	z	z_teo	p_value	overlap	f_sovrap	results
eblup.area	1.6858545	1.96	0.09182384	54	1.000000	Accept H0: The overlap is 95%

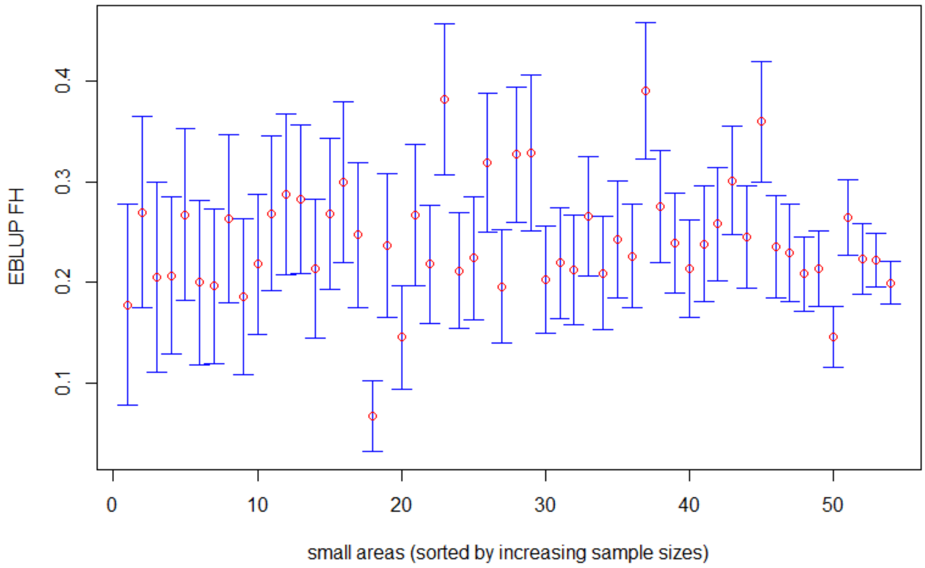

Figure 4.7.3 Confidence intervals for the Fay and Herriot estimators of the headcount ratio in Greece for the year 2013 (sorted by increasing sample sizes)

Table 4.7.5 The values of the empirical z, of the theoretical z (z_teo), the p-value, the overlapped areas, the overlap rate (f_sovrap), and the result of the test for the Fay and Herriot model estimator of the poverty gap in Greece for the year 2013

method	z	z_teo	p_value	overlap	f_sovrap	results
eblup.area	1.0614639	1.96	0.2884791	53	0.9814815	Accept H0: The overlap is 95%

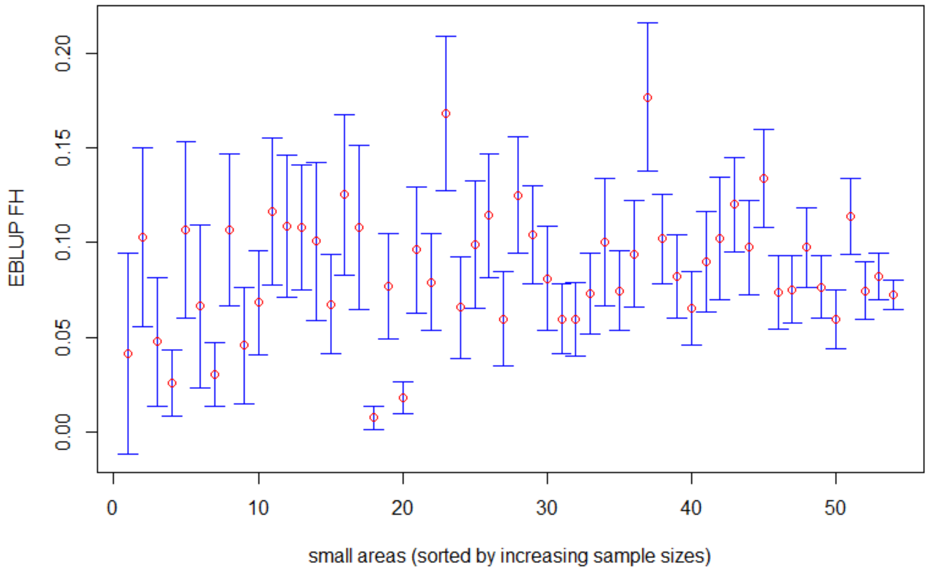

Figure 4.7.4 Confidence intervals for the Fay and Herriot estimators of the poverty gap in Greece for the year 2013 (sorted by increasing sample sizes)

iii) To test the hypothesis of normal distribution of the sampling errors a Q-Q plot for the standardized residuals, a Shapiro-Wilk test for normality, a plot of Fay-Herriot model versus standardized residuals as well as a histogram of the residuals were produced (analyzed in section 4.7 (iv))[66]. Figures 4.7.5 and 4.7.6 as well as Table 4.7.6 corresponds to the headcount ratio while Figures 4.7.7 and 4.7.8 as well as Table 4.7.7 corresponds to poverty gap.

As far as headcount ratio is concerned, the normal Q-Q plot of standardized residuals (Figure 4.7.5) shows that standardized residuals are normally distributed since they lie on a straight line, even if there are some outliers. The Shapiro-Wilk test for normality confirms the above finding since with the p-value = 0.9973 we cannot reject the null hypothesis of normality of the sampling errors. The hypothesis of normality is also confirmed from the histogram of the residuals. Furthermore, in the plot of Fay-Herriot model estimates versus standardized residuals there is not an obvious pattern in those residuals. Therefore, it seems that the assumption of constant variance of the sampling errors is satisfied,

As far as the poverty gap is concerned, although the histogram and the density plot show a slight skew, the hypothesis of normality of the sampling errors is confirmed

[66] All computations were performed using software R.

by both Q-Q plot and Shapiro-Wilk test. In the plot of Fay-Herriot model versus standardized residuals there is not an obvious pattern, so it seems that the assumption of constant variance of the sampling errors is satisfied.

Concluding, the hypothesis of normality of the sampling errors seems to be satisfied both in the case of the estimation of headcount ratio and in the case of the estimation of poverty gap.

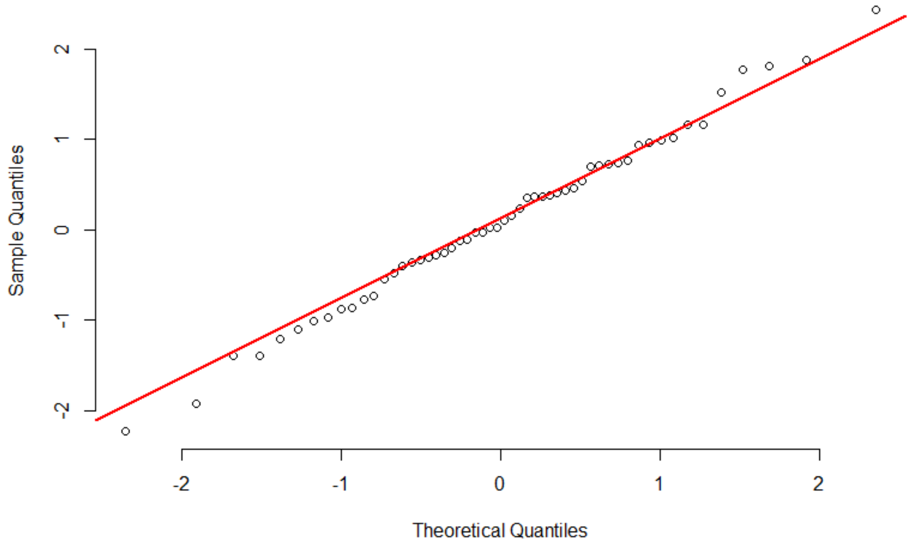

Figure 4.7.5 Normal Q-Q plot of standardized residuals of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2013

Table 4.7.6 Shapiro-Wilk test for normality of the sampling errors of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2013

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.99454	p-value=0.9973

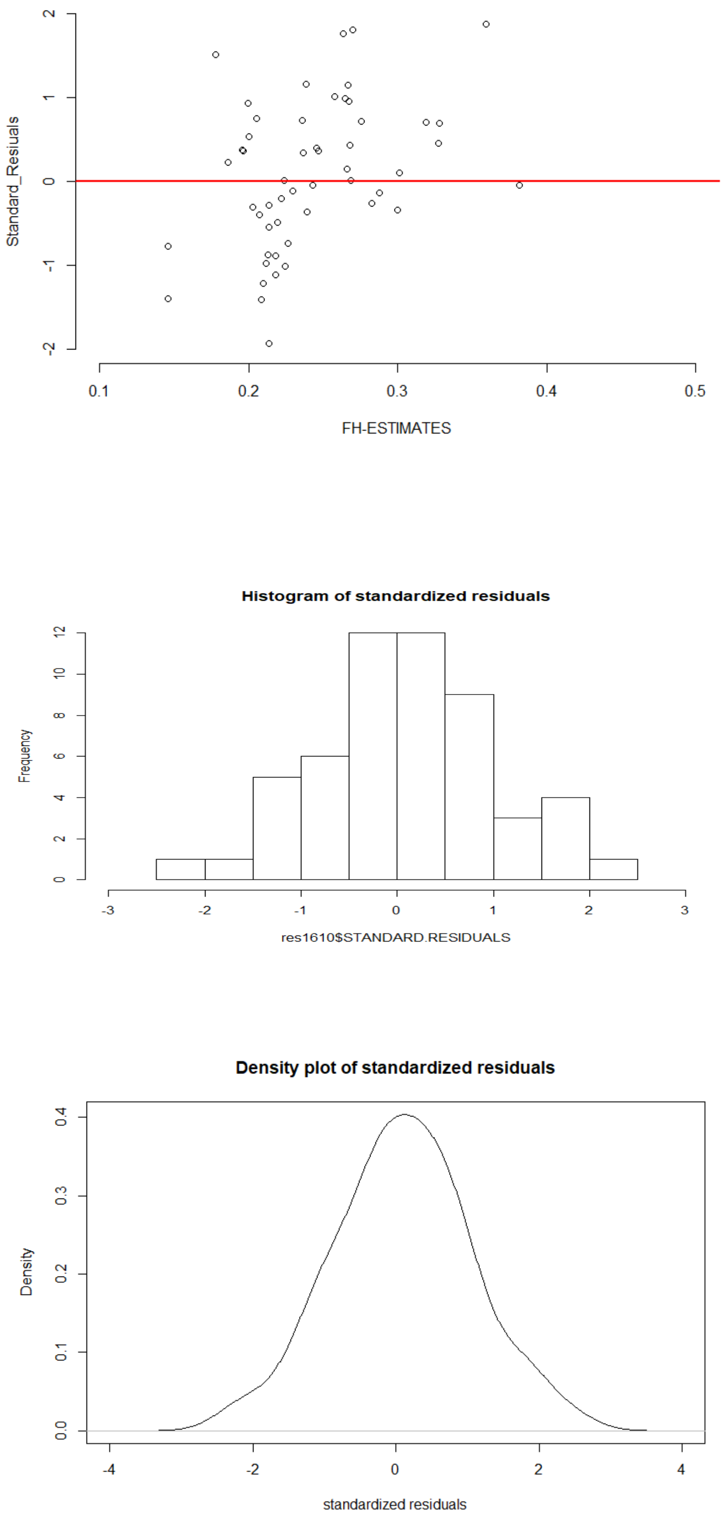

Figure 4.7.6 Residual distribution of the Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2013

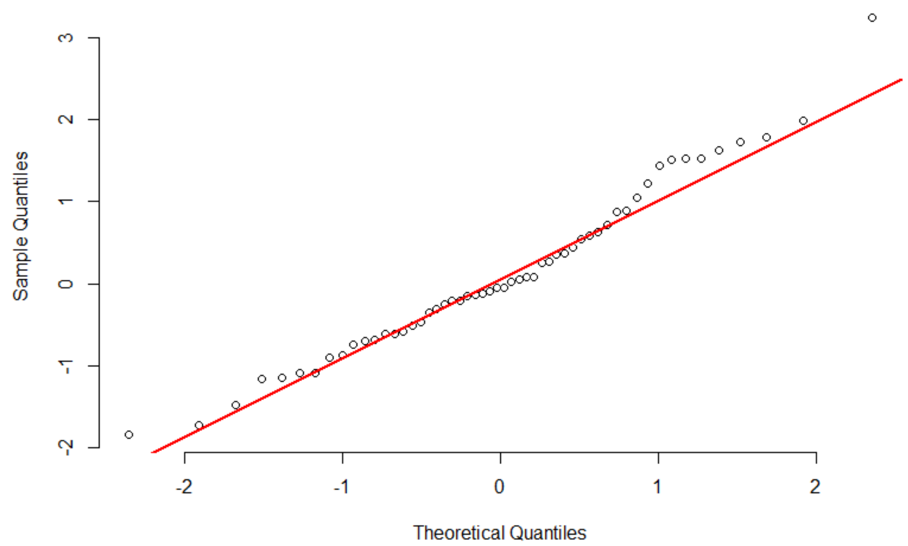

Figure 4.7.7 Normal Q-Q plot of standardized residuals of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2013

Table 4.7.7 Shapiro-Wilk test for normality of the sampling errors of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2013

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.97298	p-value=0.2599

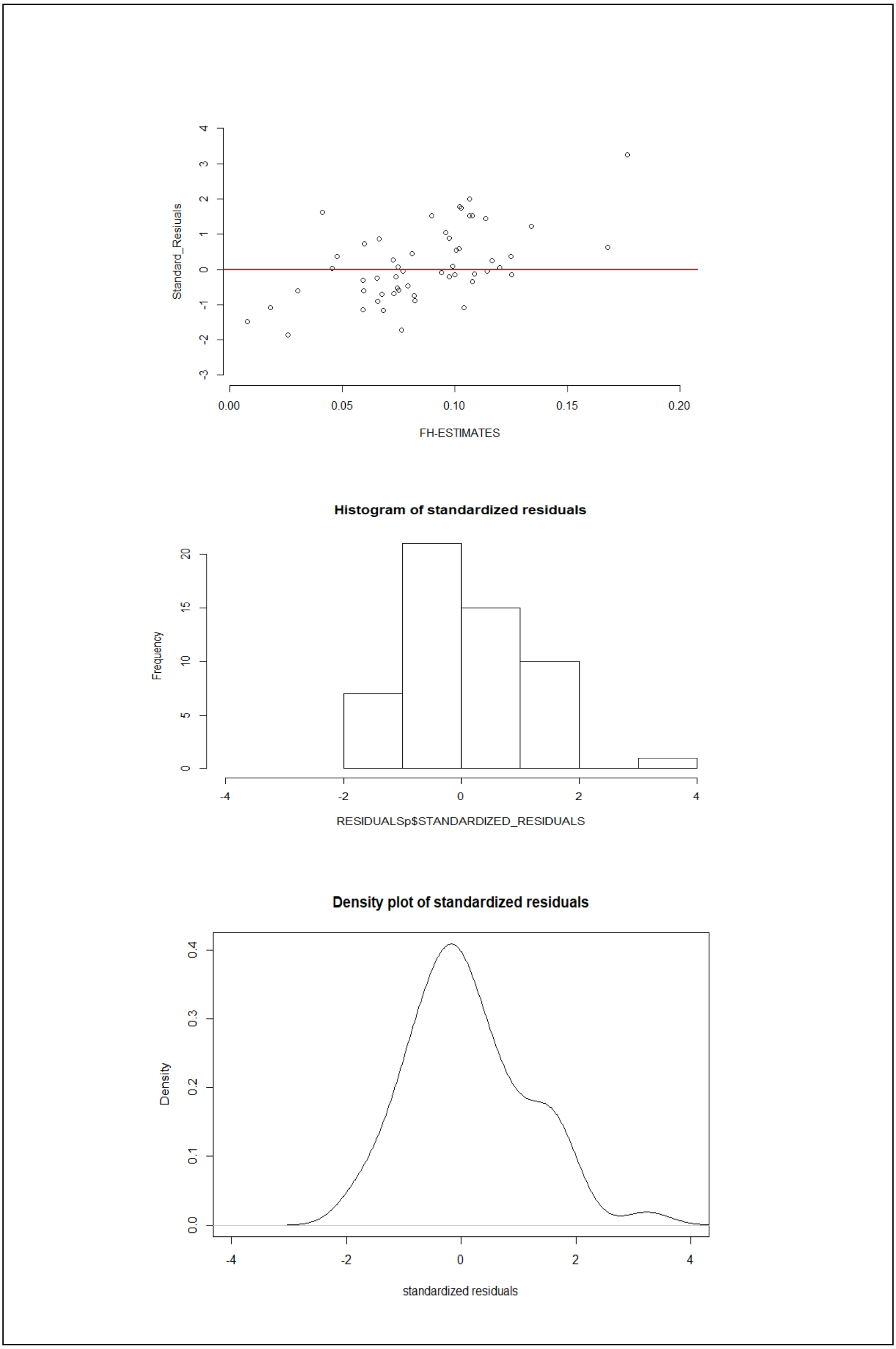

Figure 4.7.8 Residual distribution of the Fay and Herriot model for the estimation of poverty gap in Greece for the year 2013

iv. A similar procedure as for the sampling errors was followed to check the hypothesis of normality of the random effects. Figures 4.7.9, 4.7.10 and Table 4.7.8 corresponds to headcount ratio while Figures 4.7.11, 4.7.12 and Table 4.7.9 to the poverty gap.

In the case of headcount ratio there are some outliers in the Q-Q plot but the Shapiro-Wilk test as well as the histogram of random effects confirm the hypothesis of normality. The cloud of points in the plot of Fay-Herriot model versus random effects has no obvious pattern.

In the case of the poverty gap the random effects lie very satisfactorily on the straight line. The Shapiro-Wilk test as well as the histogram of random effects confirm the hypothesis of normality. Also, there is no obvious pattern in the plot of Fay-Herriot model estimates versus random effects.

In conclusion, the hypothesis of normality of random effects seems to be satisfied.

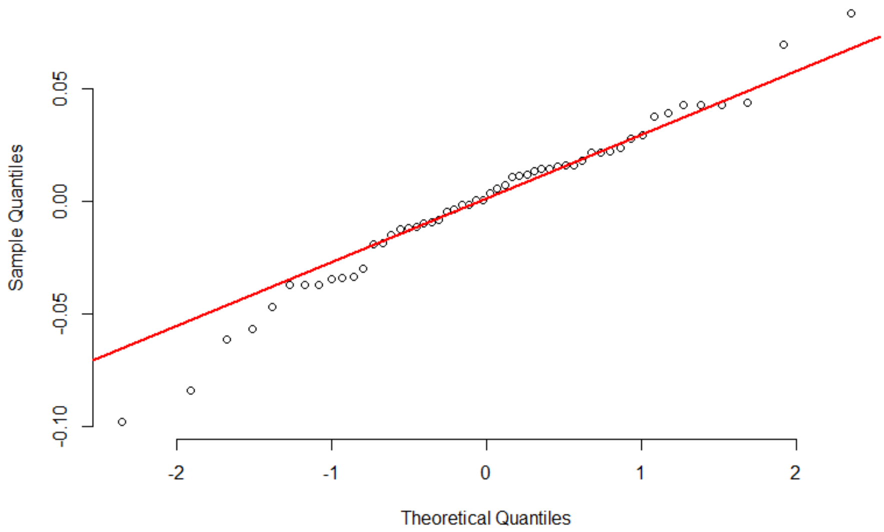

Figure 4.7.9 Normal Q-Q plot of random effects of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2013

Table 4.7.8 Shapiro-Wilk test for normality of the random effects of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2013

Shapiro-Wilk normality test	
data: random effects	
w=0.97804	p-value=0.4211

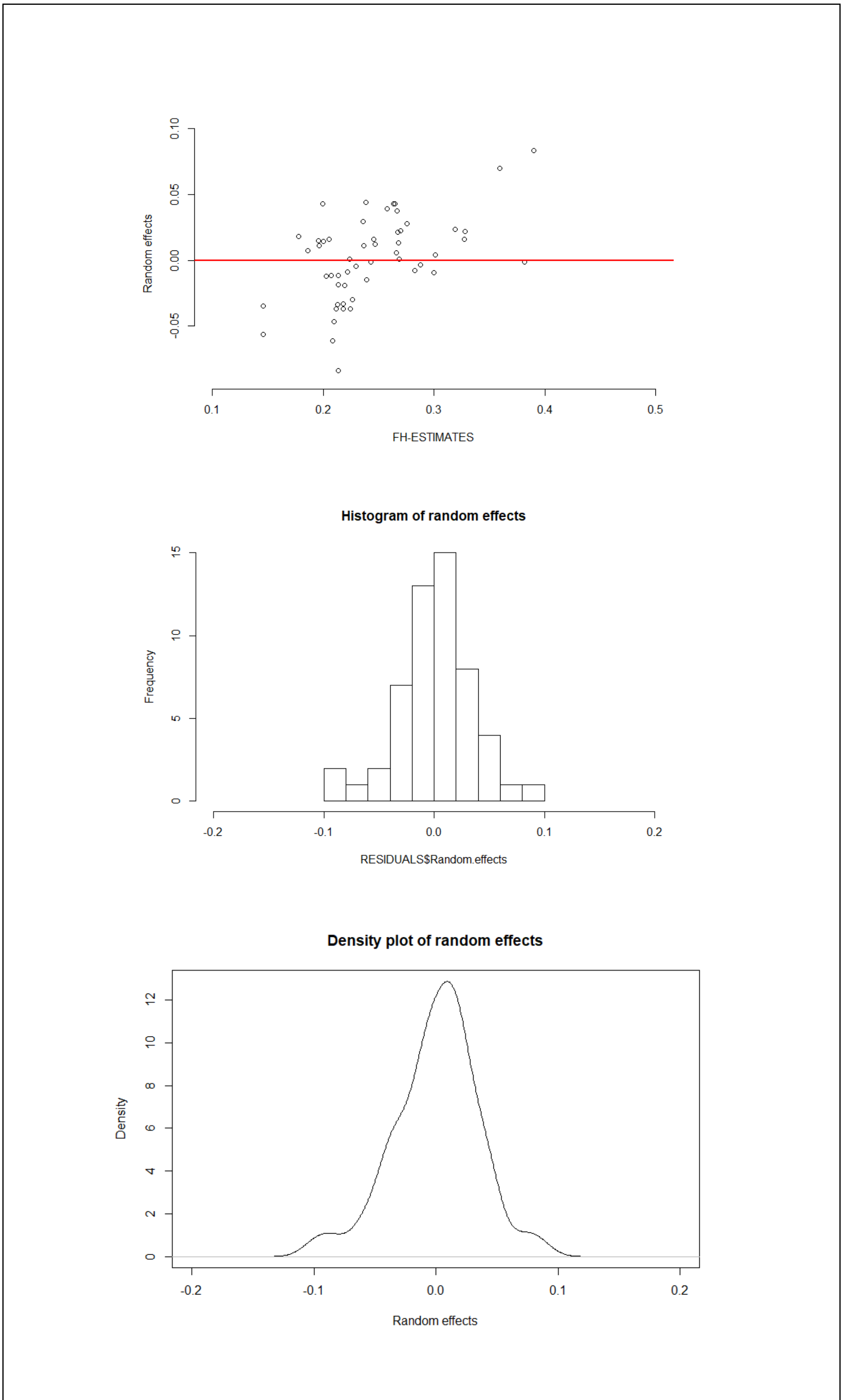

Figure 4.7.10 Random effects distribution of the Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2013

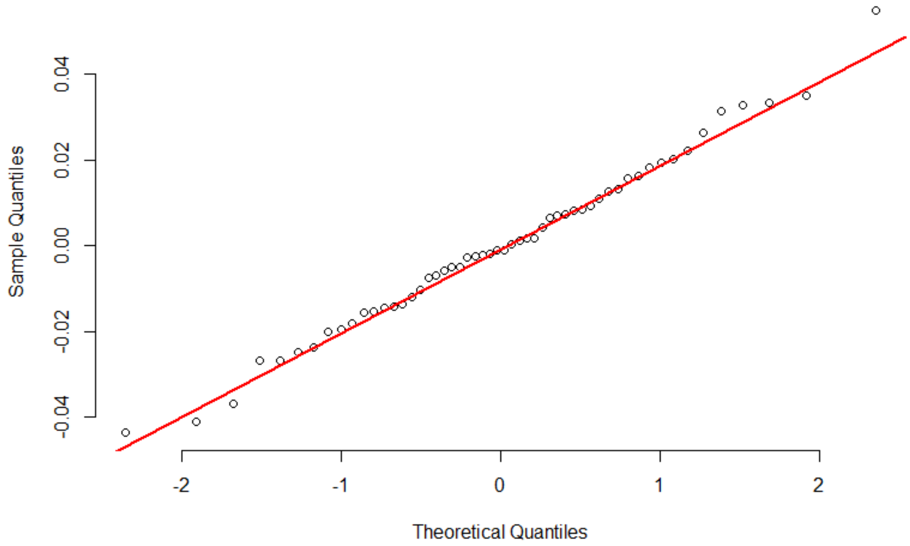

Figure 4.7.11 Normal Q-Q plot of random effects of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2013

Table 4.7.9 Shapiro-Wilk test for normality of the random effects of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2013

Shapiro-Wilk normality test	
data: random effects	
w=0.99164	p-value=0.9692

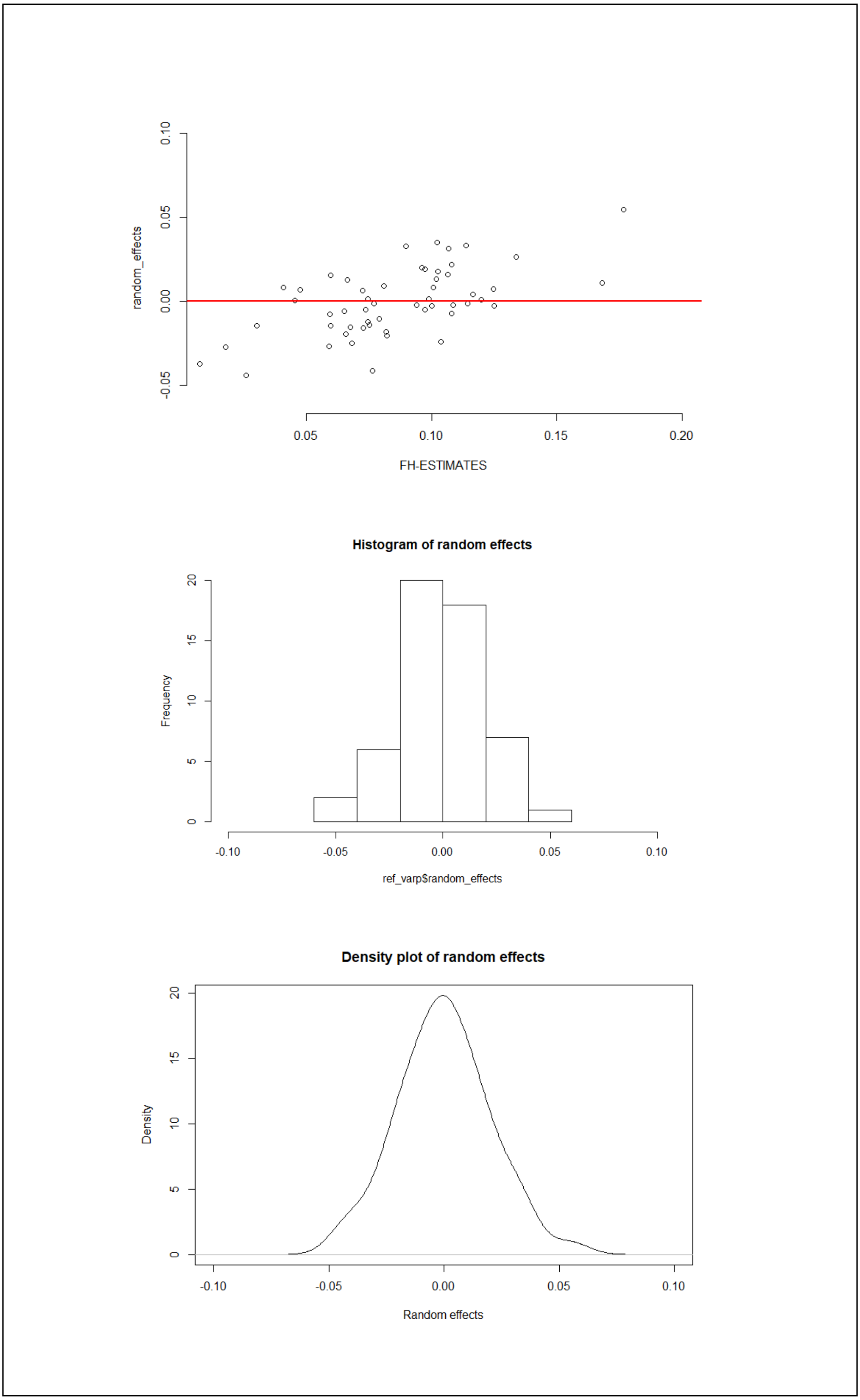

Figure 4.7.12 Random effects distribution of the Fay and Herriot model for the estimation of poverty gap in Greece for the year 2013

v) In order to evaluate the models, their quality should be checked. As already described in paragraphs 2.6.2 and 7.4 (v and vi) coefficient of variation (CV) and mean square error (MSE) have been used as quality measures. Detailed information about direct estimates, Fay and Herriot estimates and the corresponding standard errors, coefficients of variation (CV), precision gain index GIP1 and GIP2 can be found in the Appendix in Tables A11 and A12. Figures 4.7.13, 4.7.14 corresponds to the headcount ratio while Figures 4.7.15, 4.7.16 to the poverty gap.

In the case of the headcount ratio, there is one area (prefecture of Athens, code: 300101) in which the CV of the direct estimator is less than the CV of the Fay and Herriot estimator. This is something to be expected since this area has a large enough sample size (n=3873) to give an accurate direct estimate. Nevertheless, there is an overall clear gain of precision when using the Fay-Herriot estimators instead of the direct estimators. This gain is seen in both the standard error ratio (GIP2) and the estimated MSE ratio (GIP1) (all the Fay-Herriot estimates have lower MSE than the corresponding direct estimates). The improvement in precision gain tends to be greater for areas with a smaller sample size. Indeed, areas with a small sample size such as the prefectures of Thesprotia (code:300032, n=31, GIP1=3.64, GIP2=1.72), Grevena (code:300051, n=37, GIP1=2.02, GIP2=1.53) and Lassithi (code:300092, n=56, GIP1=2.04, GIP2=1.59) have a large gain in precision. For example, in Thesprotia the standard error of the Fay-Herriot estimate was reduced 3.64 times and the CV 1.72 times in relation to the direct estimate. This is evident as the direct estimator is likely to be more unstable in areas with small sample size.

In the case of the poverty gap, there is one area (prefecture of Samos, code:300084) in which the CV of the Fay-Herriot estimator is greater than the CV of the direct estimator. On the one hand this is contrary to what we expected since this area has a small sample size but on the other hand a possible explanation of the result could be the very large variance of the direct estimator (15,287%). Except for this case there is a precision gain in all other areas. Also, all the Fay-Herriot estimates have lower MSE than the corresponding direct estimates. The gain seems to be greater for areas with a smaller sample size. For example, some of them are the prefectures of Lassithi (code:300092, n=56, GIP1=2.3, GIP2=1.36), Kerkyra (code:300022, n=113, GIP1=1.57, GIP2=1.37) and Rethymno (code:300093, n=125, GIP1=1.53, GIP2=1.58).

Summarizing, the application of small area estimation approaches achieved an overall significant efficiency gain both for the estimation of the headcount ratio and for the estimation of the poverty gap.

National statistical offices usually establish a maximum publishable CV. As pointed out by Molina and Marhuenda (2015) and ONS (2004), estimates are considered sufficiently and are suitable for publication when the majority of the CV are below 20%. For the present data:

- in the case of the headcount ratio, the estimated CVs of direct estimators exceeded the level of 20% for 13 (out of the 54) domains while those of the EBLUP F-H estimators exceeded this level for only three domains.
- in the case of the poverty gap, the estimated CVs of direct estimators exceeded the level of 20% for 22 (out of the 54) domains while those of the EBLUP F-H estimators exceeded this level for nine domains.

In conclusion, the assumptions of the Fay-Herriot model seem to be satisfied for both headcount ratio and poverty gap. Also, there is a clear overall precision gain from the application of the Fay-Herriot model. The application of small area estimation approaches achieved an overall significant efficiency gain both for the estimation of the headcount ratio and for the estimation of the poverty gap. Finally, the majority of CV are below 20%, with only three exceptions for the headcount ratio and nine for the poverty gap.

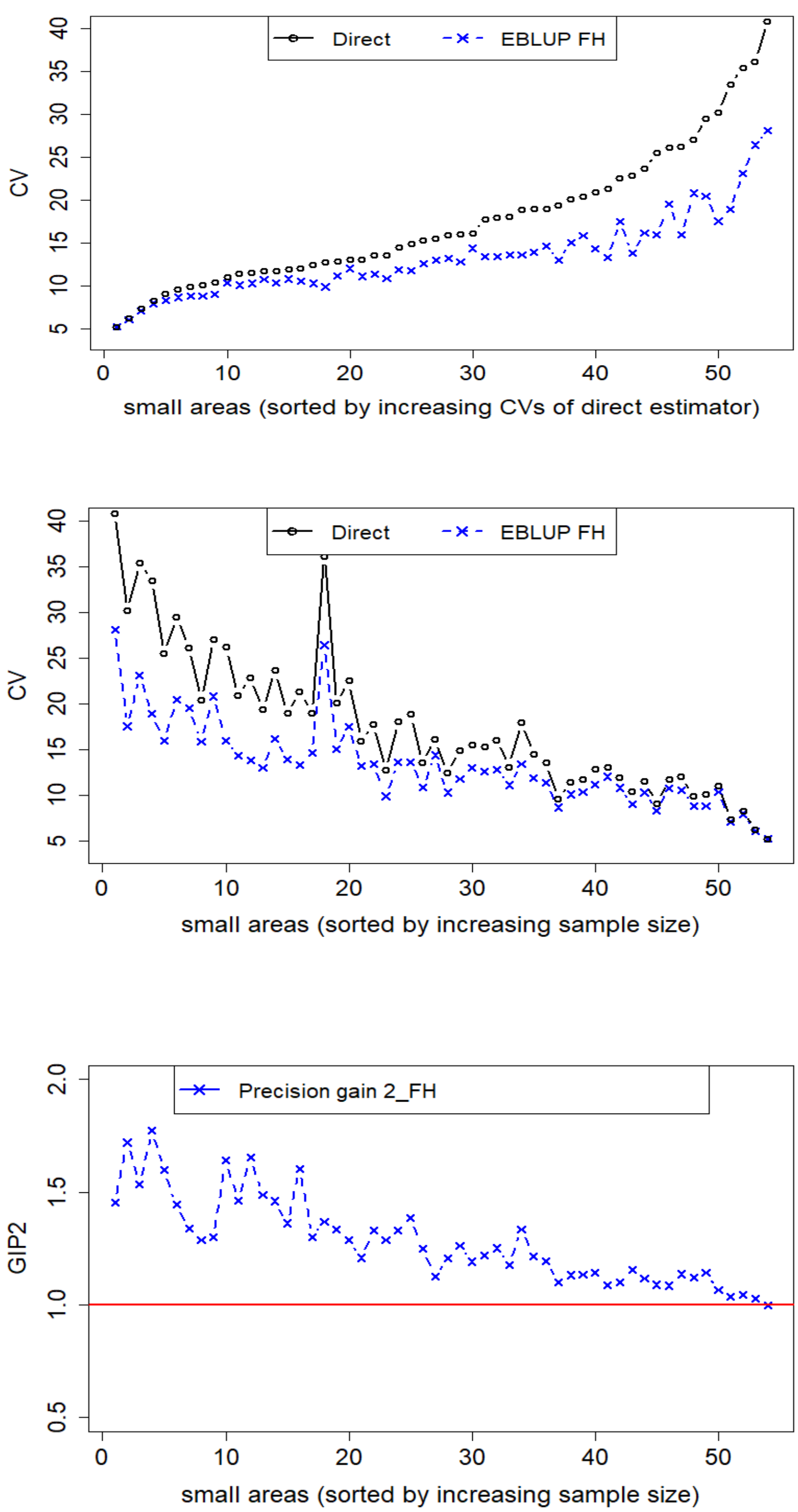

Figure 4.7.13 Coefficients of variation for the Direct and Fay-Herriot estimator of the headcount ratio in Greece for the year 2013 (sorted by increasing sample size and sorted by increasing CVs of direct estimator) and gain in precision index (GIP2)

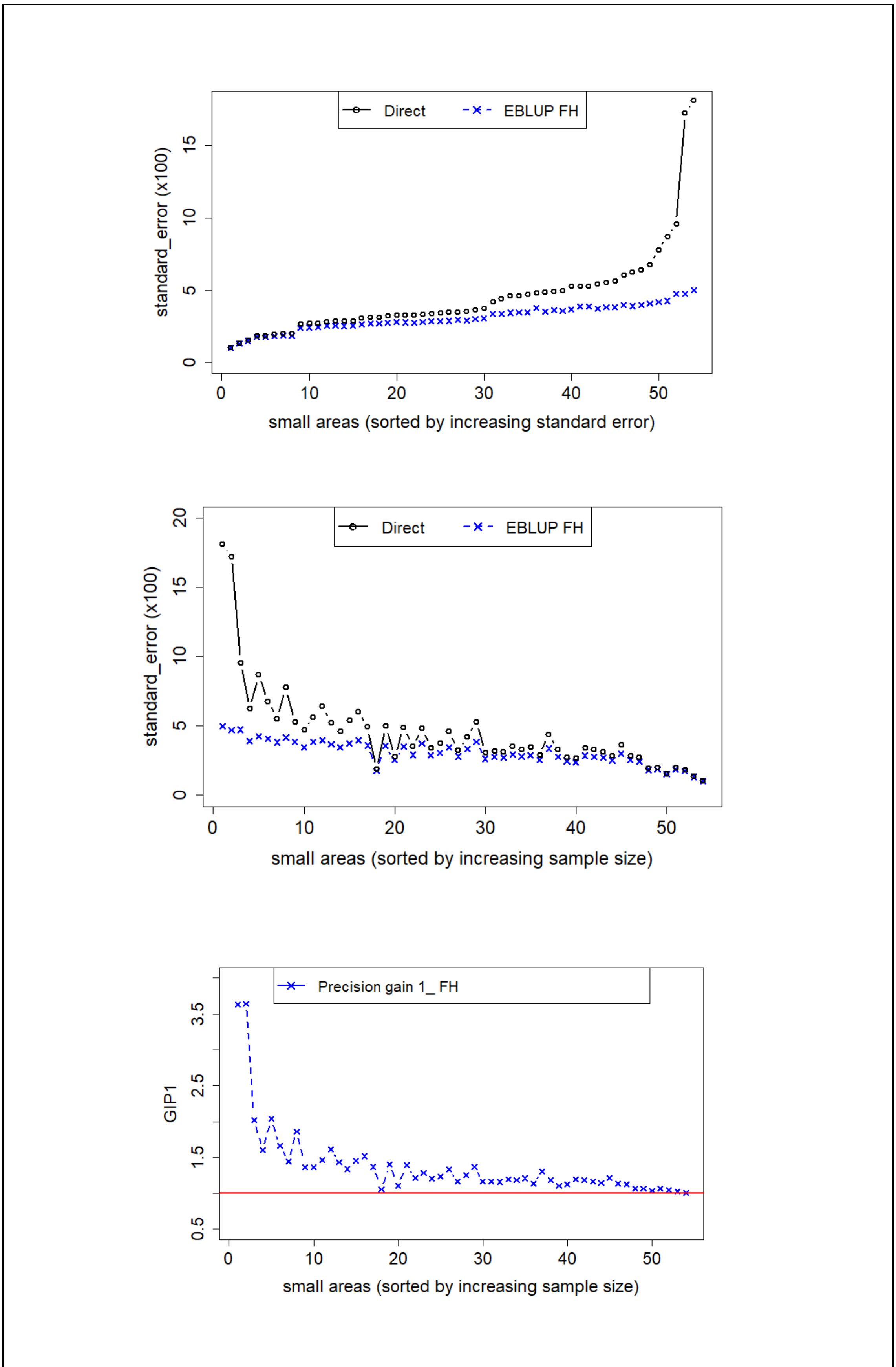

Figure 4.7.14 Standard errors for the Direct and Fay-Herriot estimator of the headcount ratio in Greece for the year 2013 and gain in precision index (GIP1) sorted by increasing sample size

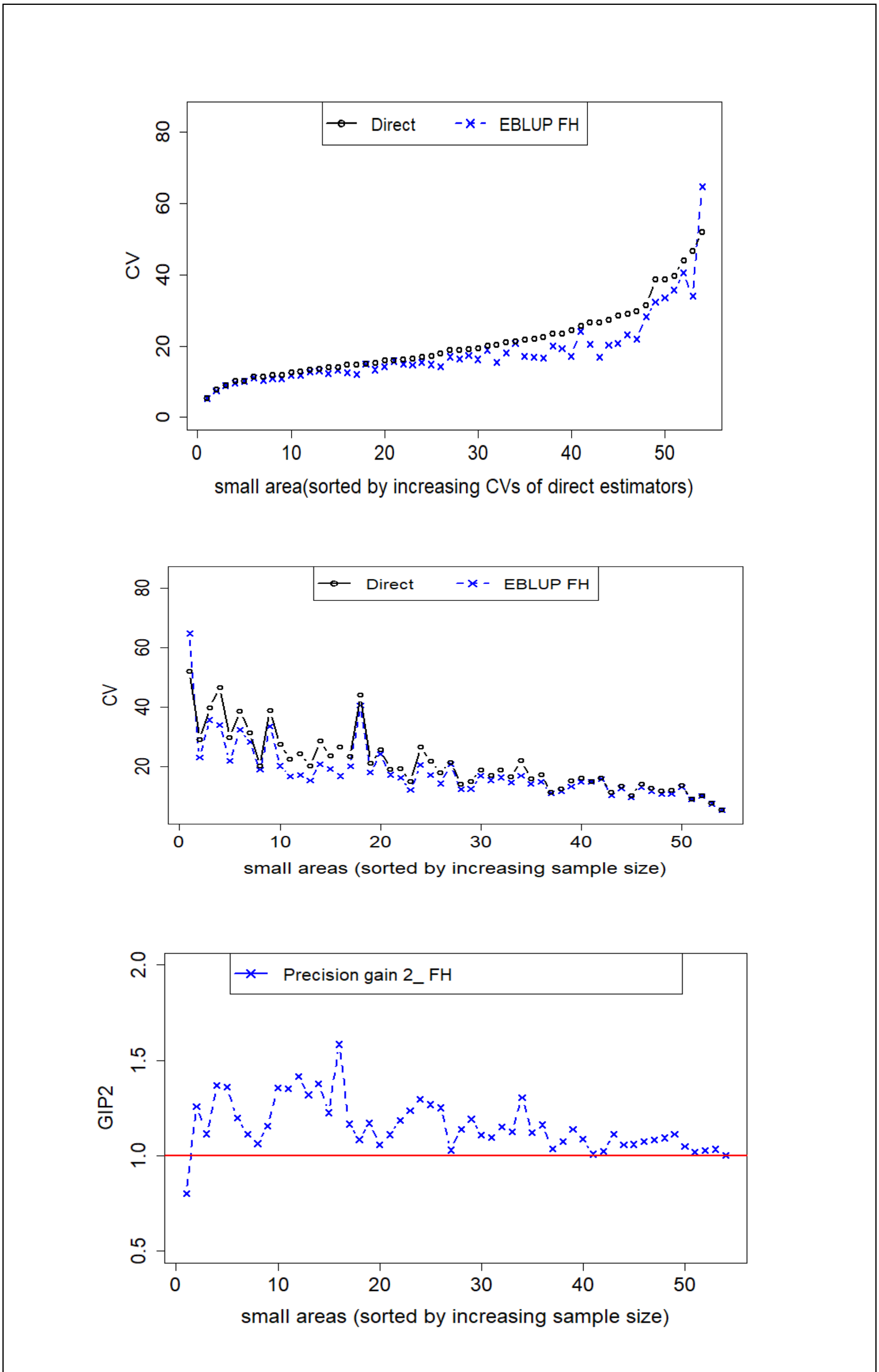

Figure 4.7.15 Coefficients of variation for the Direct and Fay-Herriot estimator of the poverty gap in Greece for the year 2013 (sorted by increasing sample size and sorted by increasing CVs of direct estimator) and gain in precision index (GIP2)

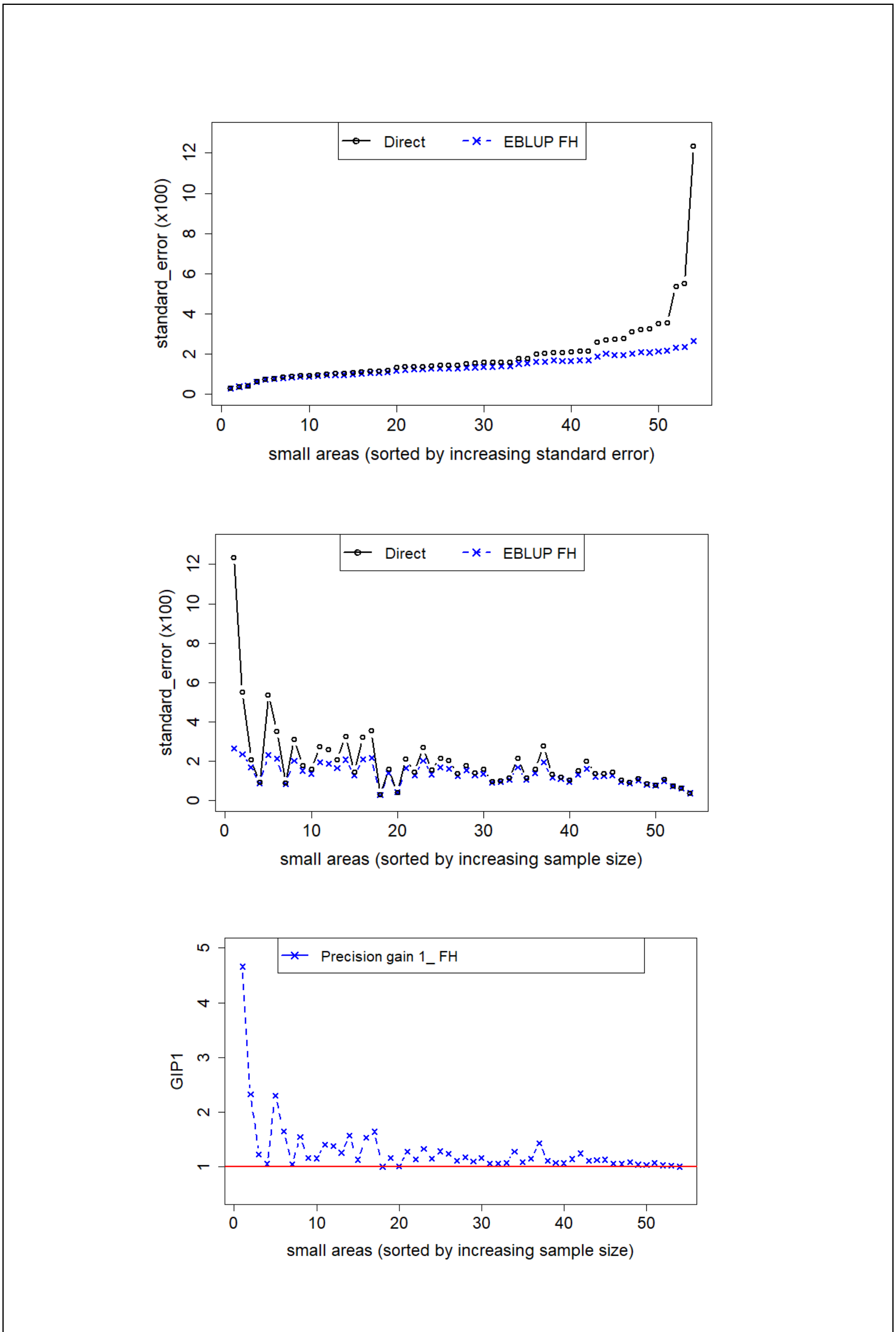

Figure 4.7.16 Standard errors for the Direct and Fay-Herriot estimator of the poverty gap in Greece for the year 2013 and gain in precision index (GIP1) sorted by increasing sample size

4.7.2 Model Diagnostics for the estimation of FGT measures in Greece for the year 2009.

For the estimation of FGT measures model 1 (Table 4.6.3) was selected. In model 1 there were two covariates: Percentage of people per prefecture with low education (X1) and percentage of inactive people per prefecture (X13). All β (beta) parameters (except intercept) have a positive sign (Table 4.7.10), which means that the increase of people with low education as well as the increase of inactive people in an area is associated with an increase of the poverty rate in this unit. The literature and EU-SILC survey confirms this finding.

Table 4.7.10 Coefficients for the final selected model for estimating the headcount ratio and poverty gap index for the year 2009 using F-H model

Headcount ratio (P_{0i})				
	beta	std.error	tvalue	pvalue
(Intercept)	-0.6445719	0.2358351	-2.733147	0.0062732340
X1 low education	0.6669005	0.1950621	3.418914	0.0006287168
X13 inactive people	0.9636103	0.4520773	2.131517	0.0330465991
Poverty gap index (P_{1i})				
	beta	std.error	tvalue	pvalue
(Intercept)	-0.1026740	0.04946243	-2.075798	0.037912591
X1 low education	0.3765594	0.11675105	3.225320	0.001258321
X13 inactive people	0.3380652	0.16666838	2.028370	0.042522457

i) In order to examine whether a substantial bias exists, the graphical diagnostic suggested by Brown et al. (2001) (analyzed in section 4.7 (i)) was produced and illustrated in Figures 4.7.17 and 4.7.18 for the estimate of headcount ratio and poverty gap respectively. In these figures the bisector is shown in black, whereas the regression line is red. Also, a goodness of fit diagnostic proposed by Brown et al. (2001) (analyzed in section 4.7 (ii)) was produced[67], and the results are given in Tables 4.7.11 and 4.7.12.

As far as the headcount ratio is concerned, the intercept of the linear regression (Figure 4.7.17) is $\beta_0 = -0.05454$ and the parameter for the slope is $\beta_1 = 1.28594$. The intercept is very close to zero, but slope estimates are slightly different from 1. There seems to be a slight disparity from the line Y=X, so there is a possible bias.

[67]For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used. This R function was developed for the application of model bias diagnostic proposed by Brown et al. (2001).

Although, the goodness of fit diagnostic, as shown in Table 4.7.11, accept null hypothesis that the Fay-Herriot estimates are close to the direct estimates.

As far as the poverty gap is concerned, the intercept of the linear regression (Figure 4.7.18) is $\beta_0 = -0.01904$ and the parameter for the slope is $\beta_1 = 1.34279$. There seems to be a disparity from the line Y=X and hence more possible bias. Although, the goodness of fit diagnostic, as shown in Table 4.7.12, accept null hypothesis that the Fay-Herriot estimates are close to the direct estimates.

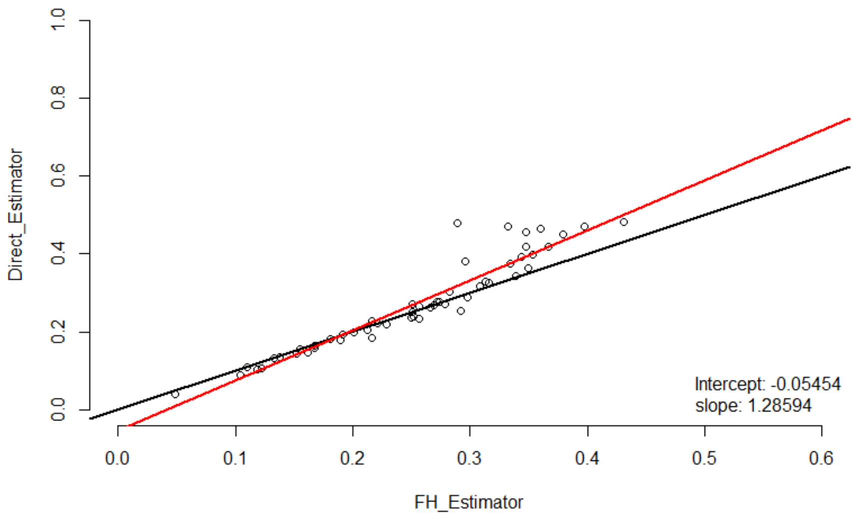

Figure 4.7.17 Relation between direct estimates and Fay-Herriot estimates of the headcount ratio in Greece for the year 2009

Table 4.7.11 The values of the empirical Wald test (W), of the theoretical χ^2 (c_alfa1), the p-value and the test result for the Fay and Herriot model estimator of the headcount ratio in Greece for the year 2009

method	W	c_alfa1	p-value	results
eblup.area	9.464634	69.83216	9.420875e-12	Accept H0: E(Direct estimates) = Model based Estimates

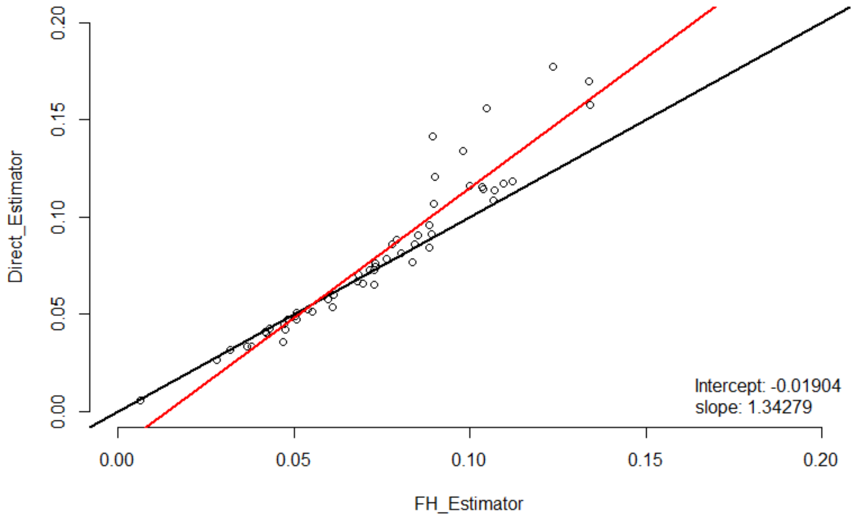

Figure 4.7.18 Relation between direct estimates and Fay-Herriot estimates of the poverty gap in Greece for the year 2009

Table 4.7.12 The values of the empirical Wald test (W), of the theoretical χ^2 (c_alfa1), the p-value and the test result for the Fay and Herriot model estimator of the poverty gap in Greece for the year 2009

method	W	c_alfa1	p-value	results
eblup.area	11.09226	69.83216	2.681184e-10	Accept H0: E(Direct estimates) = Model based Estimates

ii) In order to evaluate the validity of the confidence intervals generated by the Fay and Herriot model a coverage diagnostic was used (analyzed in section 4.7 (iii))[68]. The results are given in Tables 4.7.13 and 4.7.14 for the headcount ratio and the poverty gap respectively. Also, the numerical values of the confidence intervals are given in tables A9 and A10 in the Appendix, for the Fay and Herriot estimates of the headcount ratio and poverty gap, respectively. An illustration of the results is presented in Figures 4.7.19 and 4.7.20. In both cases the null hypothesis that the overlap is 95% is accepted. This means that the confidence intervals generated by Fay and Herriot model are valid.

[68] For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used.

Table 4.7.13 The values of the empirical z, of the theoretical z (z_teo), the p-value, the overlapped areas, the overlap rate (f_sovrap), and the result of the test for the Fay and Herriot model estimator of the headcount ratio in Greece for the year 2009

method	z	z_teo	p_value	overlap	f_sovrap	results
eblup.area	1.670172	1.96	0.09488539	53	1.000000	Accept H0: The overlap is 95%

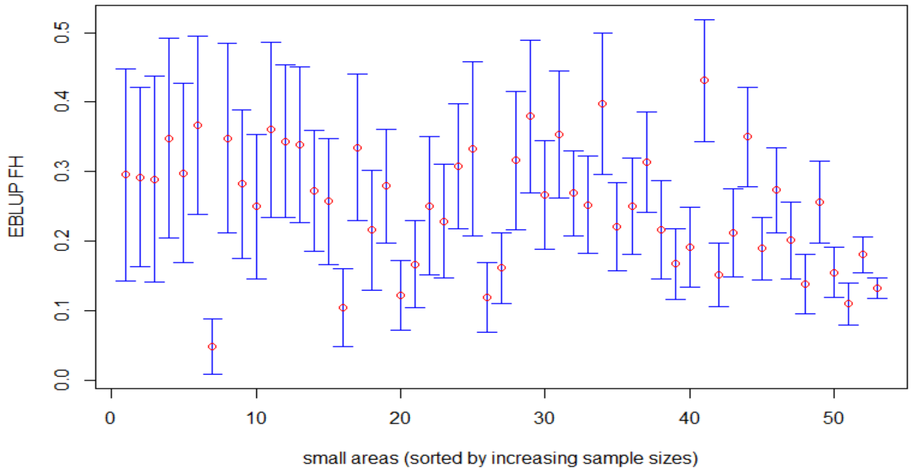

Figure 4.7.19 Confidence intervals for the Fay and Herriot estimators of the headcount ratio in Greece for the year 2009 (sorted by increasing sample sizes)

Table 4.7.14 The values of the empirical z, of the theoretical z (z_teo), the p-value, the overlapped areas, the overlap rate (f_sovrap), and the result of the test for the Fay and Herriot model estimator of the poverty gap in Greece for the year 2009

method	z	z_teo	p_value	overlap	f_sovrap	results
eblup.area	1.6701718	1.96	0.09488539	53	1.000000	Accept H0: The overlap is 95%

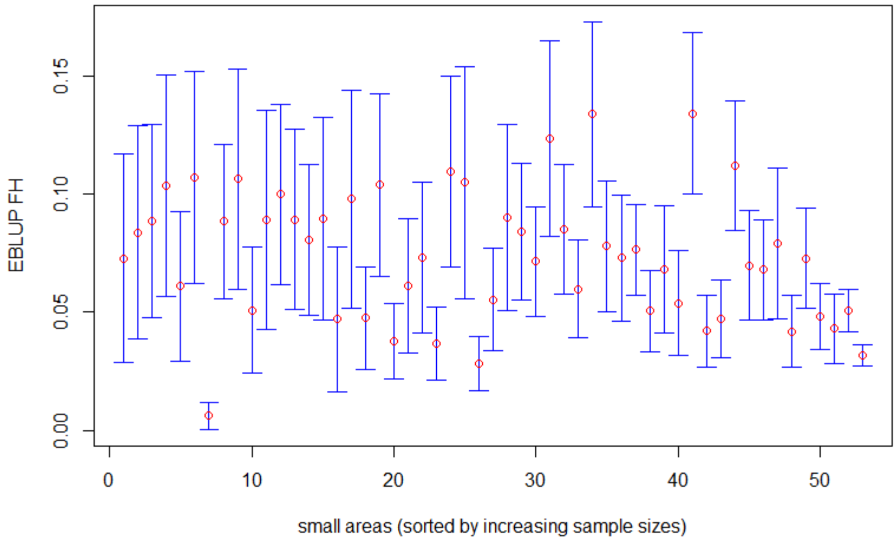

Figure 4.7.20 Confidence intervals for the Fay and Herriot estimators of the poverty gap in Greece for the year 2009 (sorted by increasing sample sizes)

iii) To test the hypothesis of normal distribution of the sampling errors a Q-Q plot for the standardized residuals, a Shapiro-Wilk test for normality, a plot of Fay-Herriot model versus standardized residuals as well as a histogram of the residuals were produced (analyzed in section 4.7 (iv))[69]. Figures 4.7.21 and 4.7.22 as well as Table 4.7.15 corresponds to the headcount ratio while Figures 4.7.23 and 4.7.24 as well as Table 4.7.16 corresponds to the poverty gap.

As far as the headcount ratio is concerned, the normal Q-Q plot of standardized residuals (Figure 4.7.21) shows that standardized residuals are normally distributed since they lie on a straight line, even if there are some outliers. The Shapiro-Wilk test for normality confirms the above finding since with the p-value = 0.3925 we cannot reject the null hypothesis of normality of the sampling errors. The above conclusions are confirmed by the histogram of the residuals where an approximately normal distribution is observed. Furthermore, in the plot of Fay-Herriot model estimates versus standardized residuals there is not an obvious pattern in those residuals. Therefore, it seems that the assumption of constant variance of the sampling errors is satisfied.

As far as the poverty gap is concerned, although there is a slight deviation in the Q-Q plot and a slight skew in the histogram and the density plot of residuals, the distribution seems not to deviate very much for normal. The hypothesis of normality of

[69] All computations were performed using software R.

the sampling errors is confirmed by Shapiro-Wilk test (p-value=0.8658). In the plot of Fay-Herriot model versus standardized residuals seems that there is a notable but not so clear pattern (a connection between the residuals and Fay-Herriot estimates) so it seems that the assumption of constant variance of the sampling errors is not satisfied. This will have an impact on the calculation of the confidence intervals.

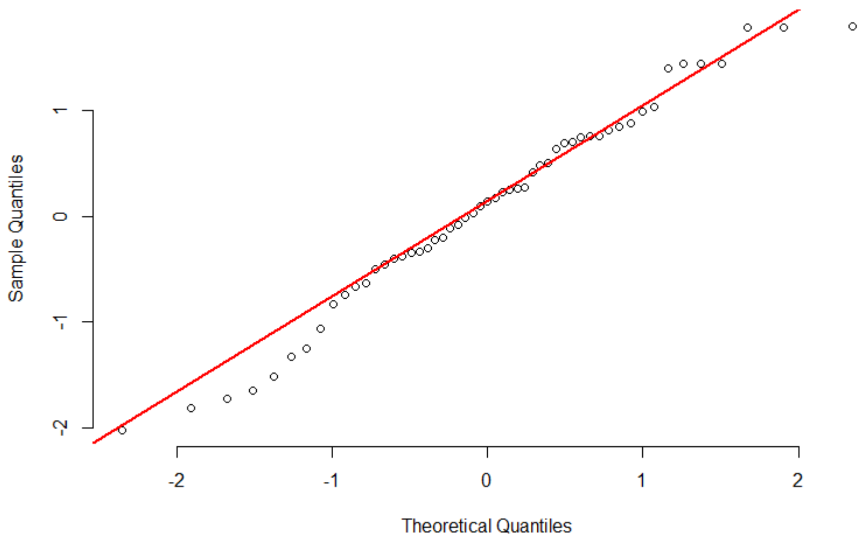

Figure 4.7.21 Normal Q-Q plot of standardized residuals of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2009

Table 4.7.15 Shapiro-Wilk test for normality of the sampling errors of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2009

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.97693	p-value=0.3925

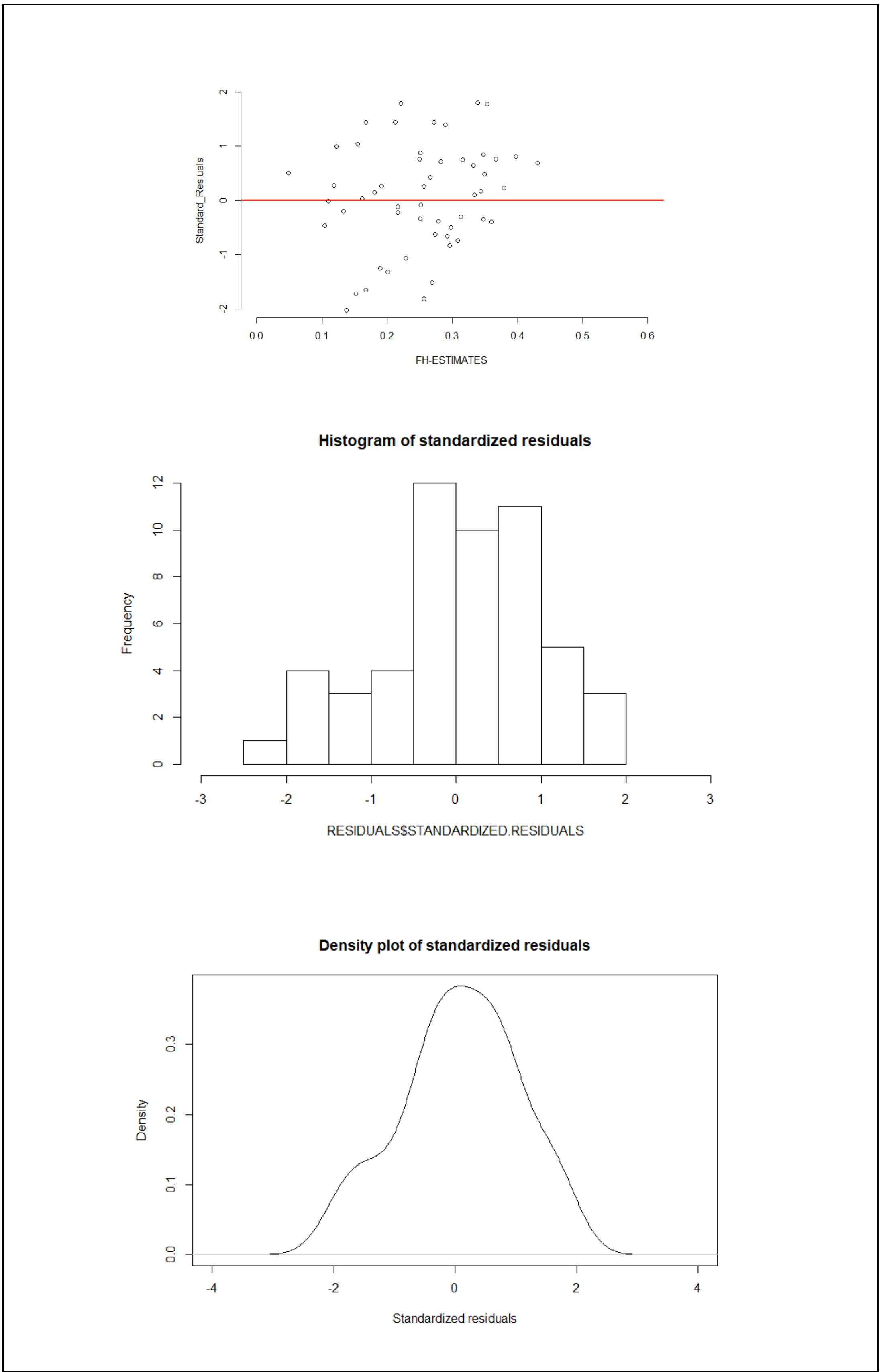

Figure 4.7.22 Residual distribution of the Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2009

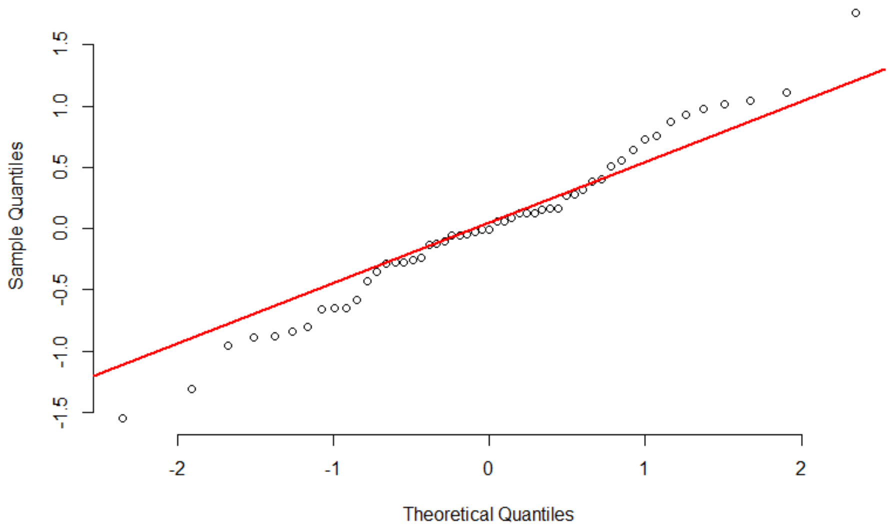

Figure 4.7.23 Normal Q-Q plot of standardized residuals of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2009

Table 4.7.16 Shapiro-Wilk test for normality of the sampling errors of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2009

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.98789	p-value=0.8658

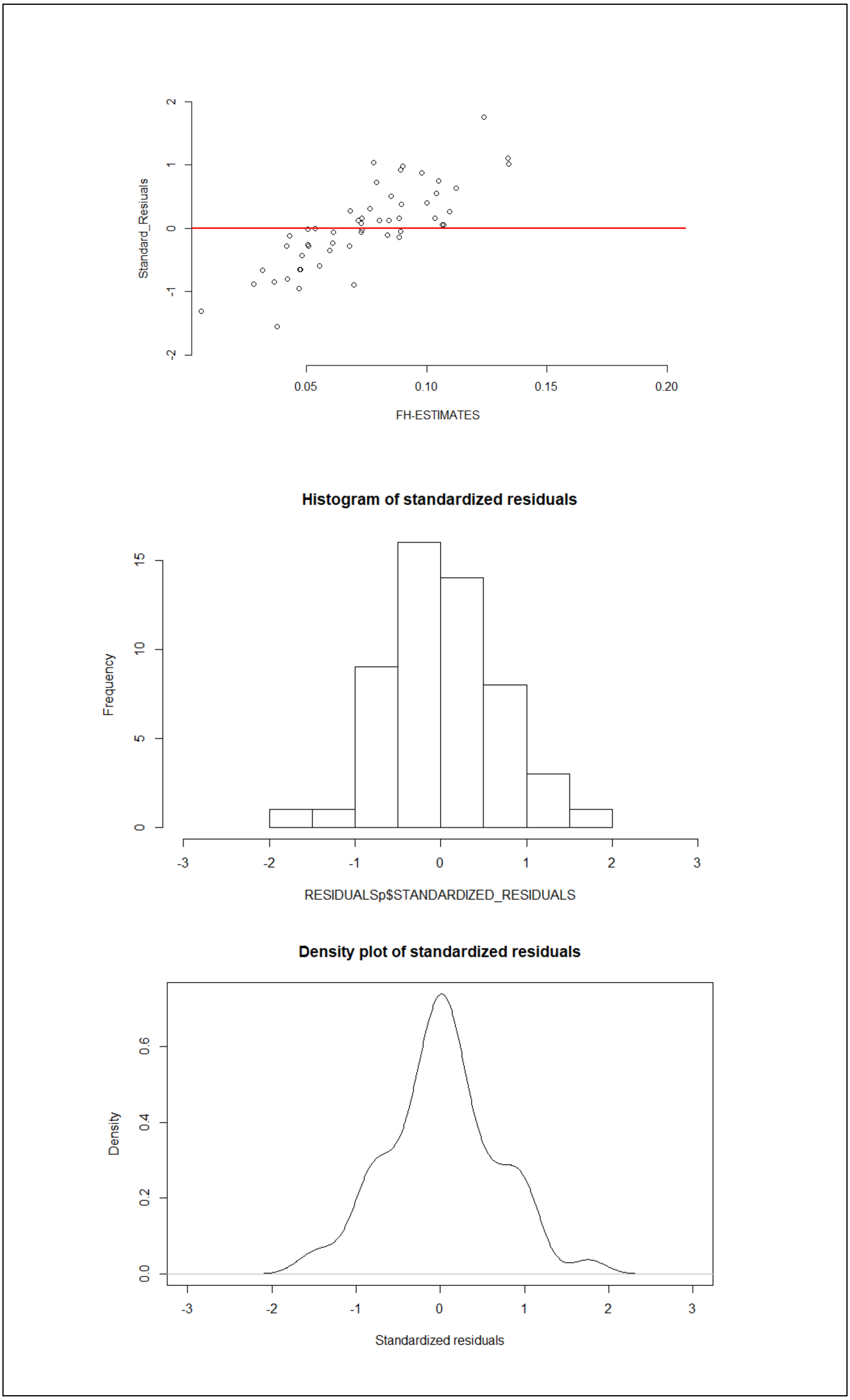

Figure 4.7.24 Residual distribution of the Fay and Herriot model for the estimation of poverty gap in Greece for the year 2009

iv) A similar procedure as for the sampling errors was followed to check the hypothesis of normality of the random effects. Figures 4.7.25, 4.7.26 and Table 4.7.17 correspond to headcount ratio while Figures 4.7.27, 4.7.28 and Table 4.7.18 to the poverty gap.

In the case of the headcount ratio there is a notable deviation from the straight line in the normal Q-Q plot as wells as a slight skew in density plot of random effects, but the Shapiro-Wilk test confirm the hypothesis of normality (p-value=0.1616). The cloud of points in the plot of Fay-Herriot model versus random effects has no obvious pattern.

In the case of the poverty gap the random effects lie quite satisfactory on the straight line. The Shapiro-Wilk test as well as the histogram of random effects confirm the hypothesis of normality. The cloud of points in the plot of Fay-Herriot model estimates versus random effects it seems to have a notable pattern.

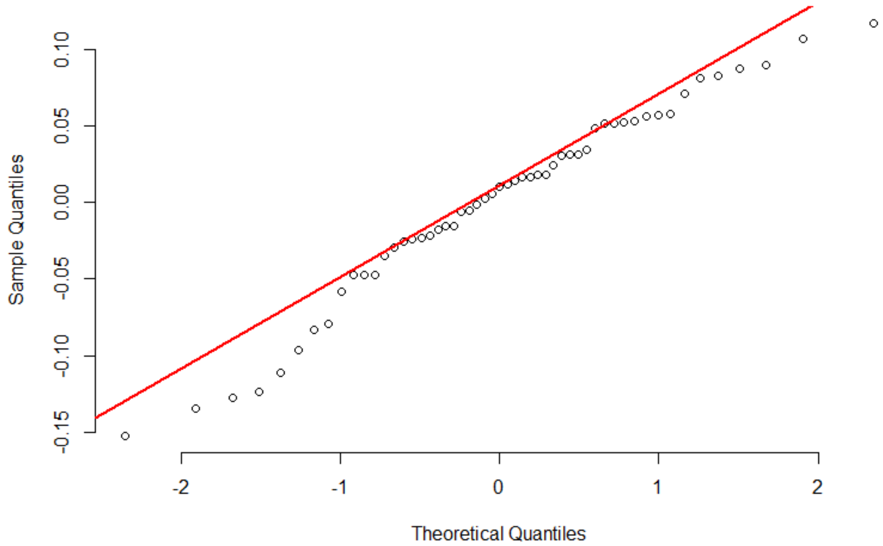

Figure 4.7.25 Normal Q-Q plot of random effects of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2009

Table 4.7.17 Shapiro-Wilk test for normality of the random effects of Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2009

Shapiro-Wilk normality test	
data: random effects	
w=0.96778	p-value=0.1616

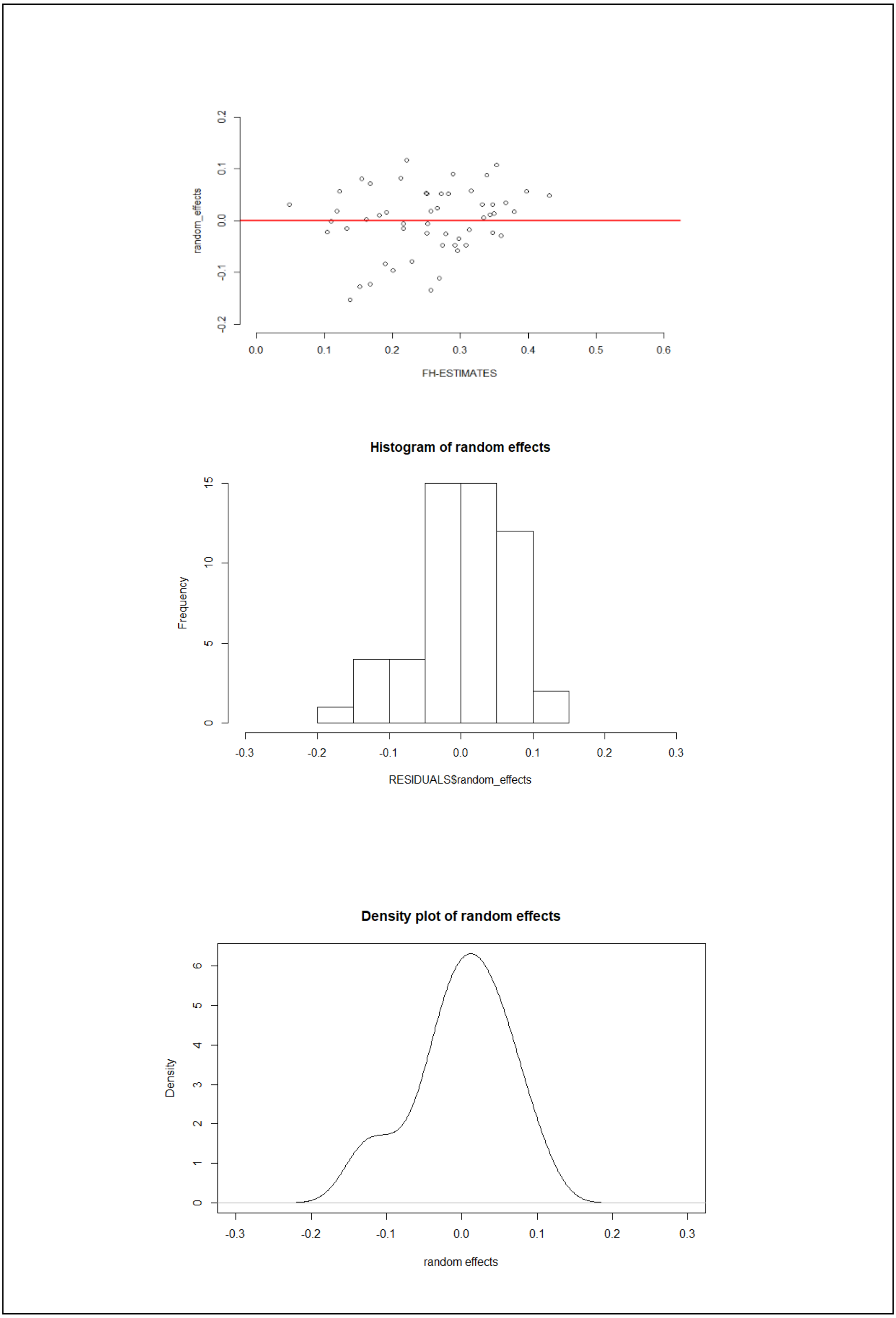

Figure 4.7.26 Random effects distribution of the Fay and Herriot model for the estimation of headcount ratio in Greece for the year 2009

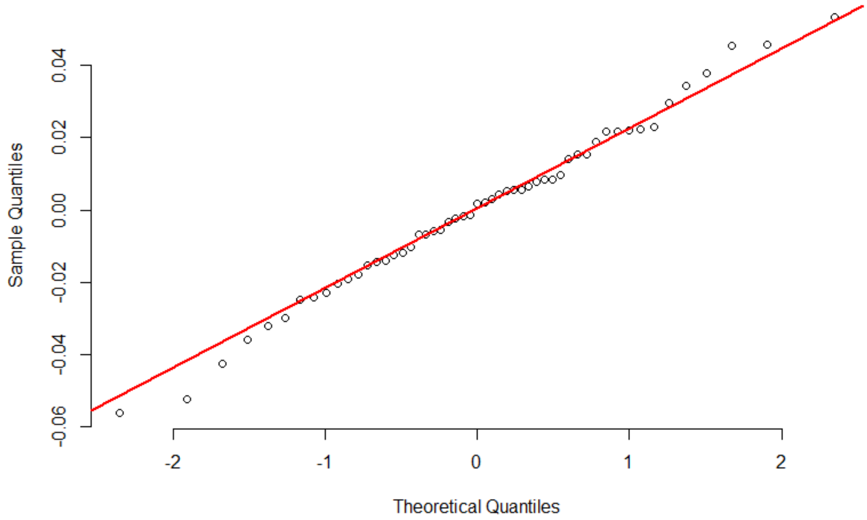

Figure 4.7.27 Normal Q-Q plot of random effects of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2009

Table 4.7.18 Shapiro-Wilk test for normality of the random effects of Fay and Herriot model for the estimation of poverty gap in Greece for the year 2009

Shapiro-Wilk normality test	
data: random effects	
w=0.99146	p-value=0.968

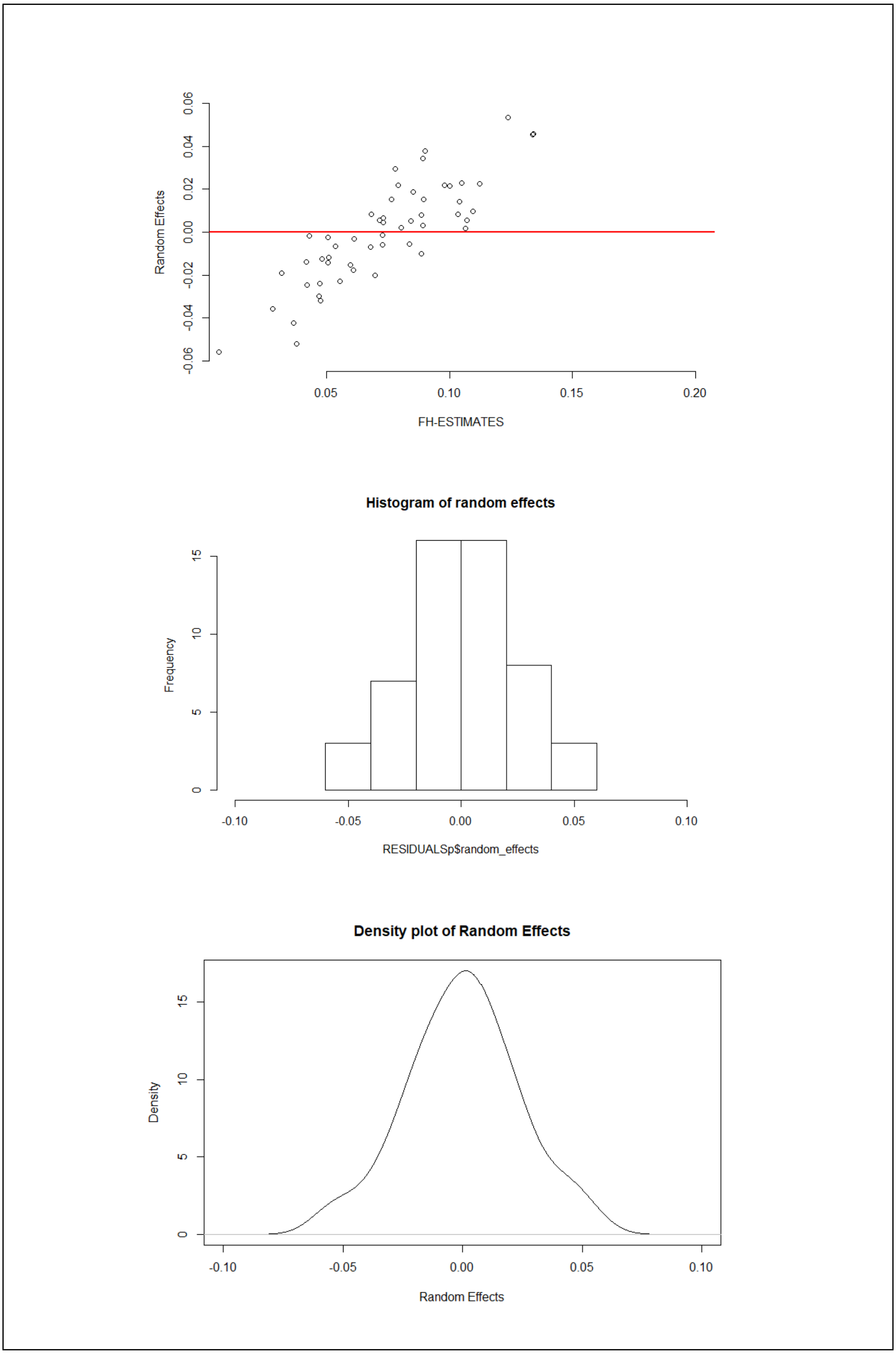

Figure 4.7.28 Random effects distribution of the Fay and Herriot model for the estimation of poverty gap in Greece for the year 2009

v) In order to evaluate the models, their quality should be checked. Coefficient of variation (CV) and mean square error (MSE) have been used as quality measures (analyzed in paragraphs 2.6.2 and 7.4 (v) and (vi)). Detailed information about direct estimates, Fay and Herriot estimates and corresponding standard errors, coefficients of variation (CV), precision gain index GIP1 and GIP2 can be found in the Appendix in Tables A13 and A14. Figures 4.7.29, 4.7.30 corresponds to the headcount ratio while Figures 4.7.31, 4.7.32 to the poverty gap.

In the case of the headcount ratio, there is one area (prefecture of Athens, code: 300101) in which the CV of the direct estimator is less than the CV of the Fay and Herriot estimator. This is something to be expected since this area has a large enough sample size (n=3934) to give an accurate direct estimate. Nevertheless, there is an overall clear gain of precision when using the Fay-Herriot estimators instead of the direct estimators. This gain is seen in both the standard error ratio (GIP2) and the estimated MSE ratio (GIP1) (all the Fay-Herriot estimates have lower MSE than the corresponding direct estimates). The improvement in precision gain tends to be greater for areas with a smaller sample size. Indeed, areas with a small sample size such as the prefectures of Evrytania (code:300005, n=35, GIP1=2.02, GIP2=1.54), Lefkada (code:300024, n=28, GIP1=1.58, GIP2=1.82), Thesprotia (code:300032, n=59, GIP1=1.6, GIP2=1,65), and Samos (code:300084, n=22, GIP1=2.49, GIP2=1,93), have a large gain in precision. For example, in Evrytania the standard error of the Fay-Herriot estimate was reduced 2.02 times and the CV 1.54 times in relation to the direct estimate. This is evident as the direct estimator is likely to be more unstable in areas with small sample size.

In the case of the poverty gap, although there are three areas (300004, 300072, 300073, 300103) where the CV of the Fay-Herriot estimator is greater than the CV of the direct estimator there is an overall gain in precision. All the Fay-Herriot estimates have lower MSE than the corresponding direct estimates. The gain seems to be greater for areas with a smaller sample size. For example, some of them are the prefectures of Evrytania (code:300005, n=35, GIP1=1.46, GIP2=1.31), Lefkada (code:300024, n=28, GIP1=1.43, GIP2=1.56) and Samos (code:300084, n=22, GIP1=1.47, GIP2=1.64).

Summarizing, the application of small area estimation approaches achieved an overall significant efficiency gain both for the estimation of the headcount ratio and for the estimation of the poverty gap.

National statistical offices usually establish a maximum publishable CV. As pointed out by Molina and Marhuenda (2015) and ONS (2004), estimates are considered sufficiently and are suitable for publication when the majority of the CV are below 20%. For the present data:

- in the case of the headcount ratio, the estimated CVs of direct estimators exceeded the level of 20% for 19 (out of the 53) domains while those of the EBLUP F-H estimators exceeded this level for seven domains.
- in the case of the poverty gap, the estimated CVs of direct estimators exceeded the level of 20% for 30 (out of the 53) domains while those of the EBLUP F-H estimators exceeded this level for 19 domains.

In conclusion, as far as headcount ratio is concerned the assumptions of Fay-Herriot model seem to be satisfied. In the case of poverty gap, the only assumption that is not clear satisfied is that of the constant variance of the sampling errors. Also, there is a clear overall precision gain from the application of the Fay-Herriot model for both of the headcount ratio and poverty gap. Finally, the majority of CV of headcount ratio are below 20% (seven exceptions) while for the poverty gap, 19 domains have greater CV of 20%.

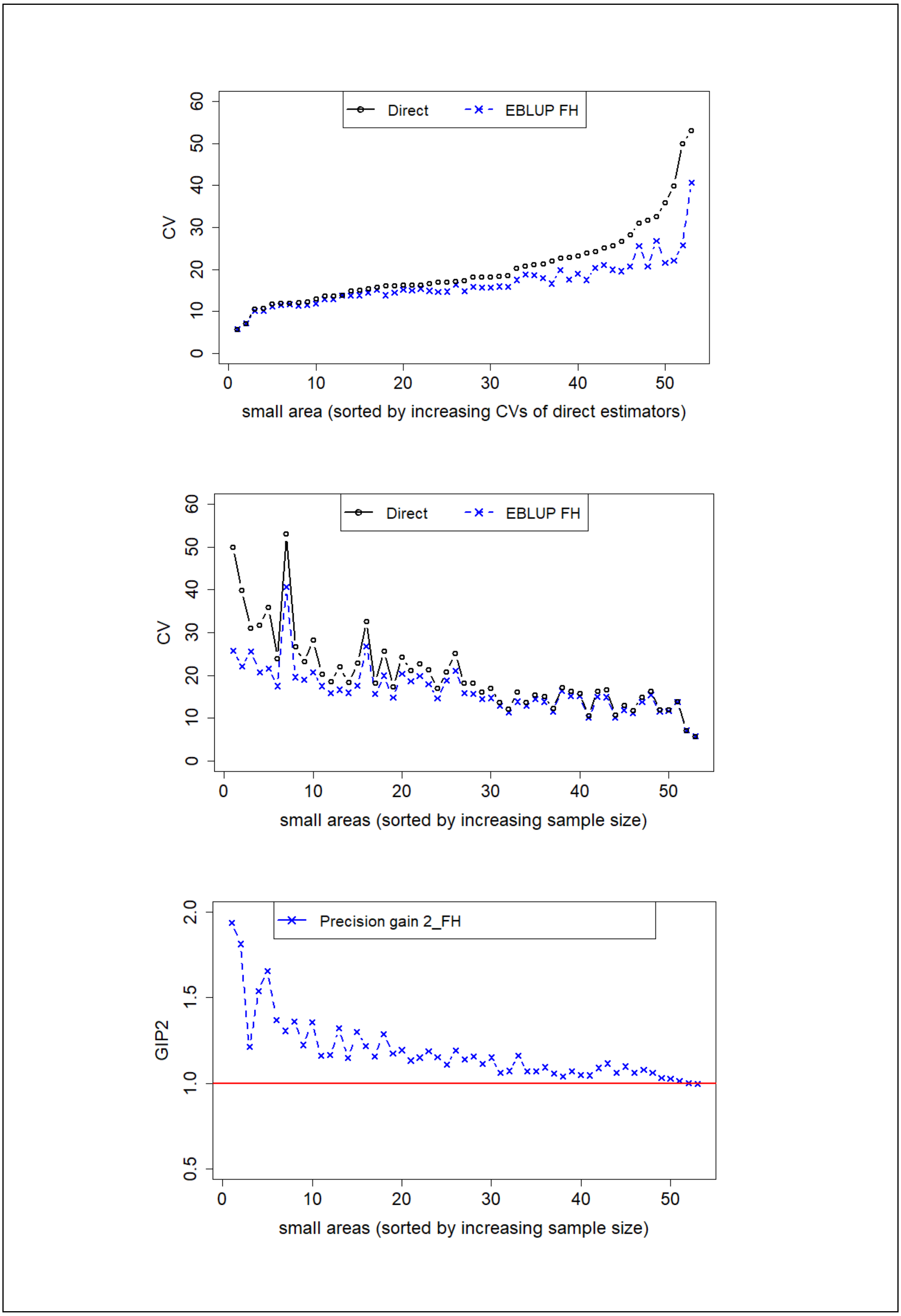

Figure 4.7.29 Coefficients of variation for the Direct and Fay-Herriot estimator of the headcount ratio in Greece for the year 2009 (sorted by increasing sample size and sorted by increasing CVs of direct estimator) and gain in precision index (GIP2)

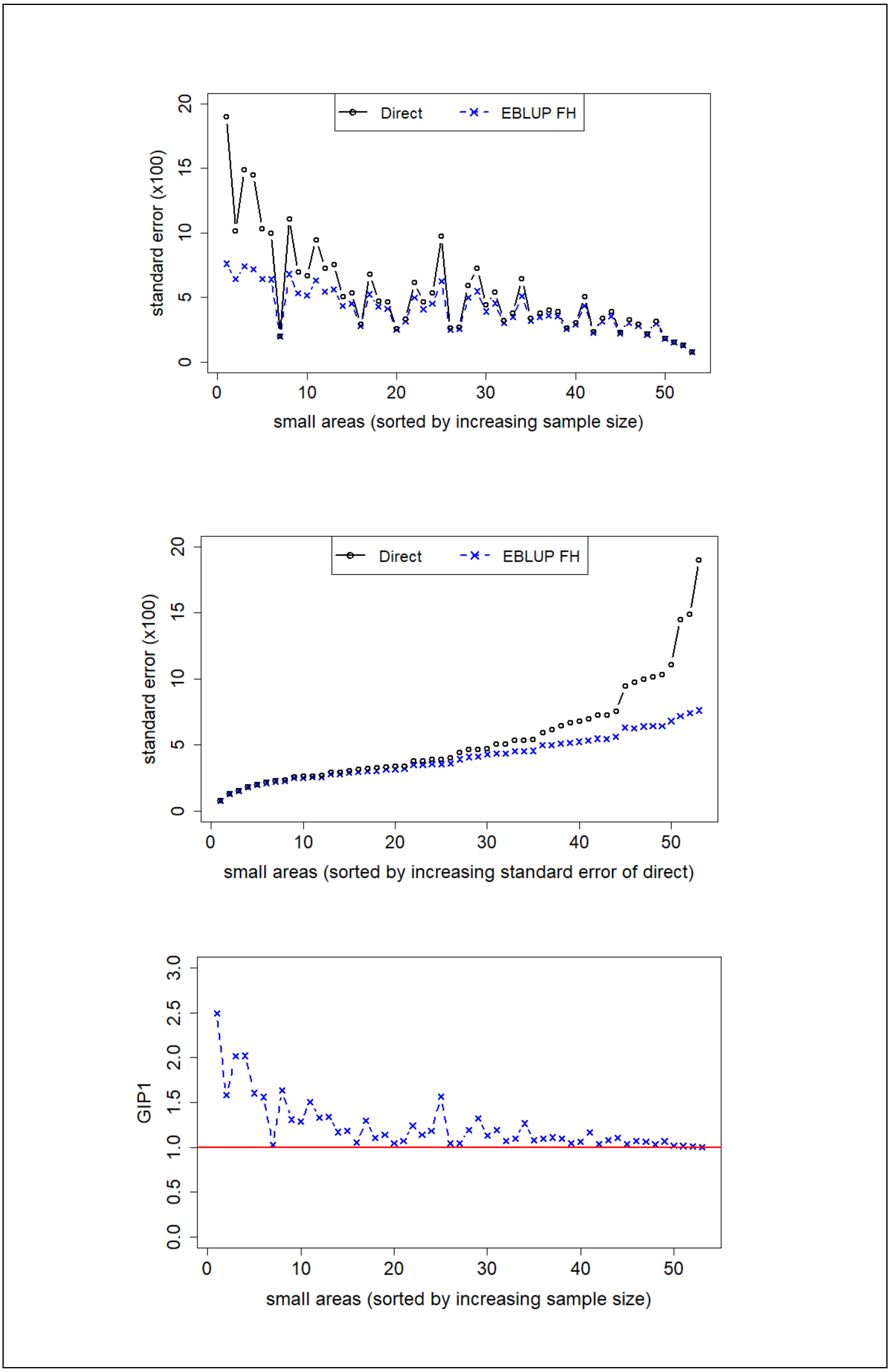

Figure 4.7.30 Standard errors for the Direct and Fay-Herriot estimator of the headcount ratio in Greece for the year 2009 and gain in precision index (GIP1) sorted by increasing sample size

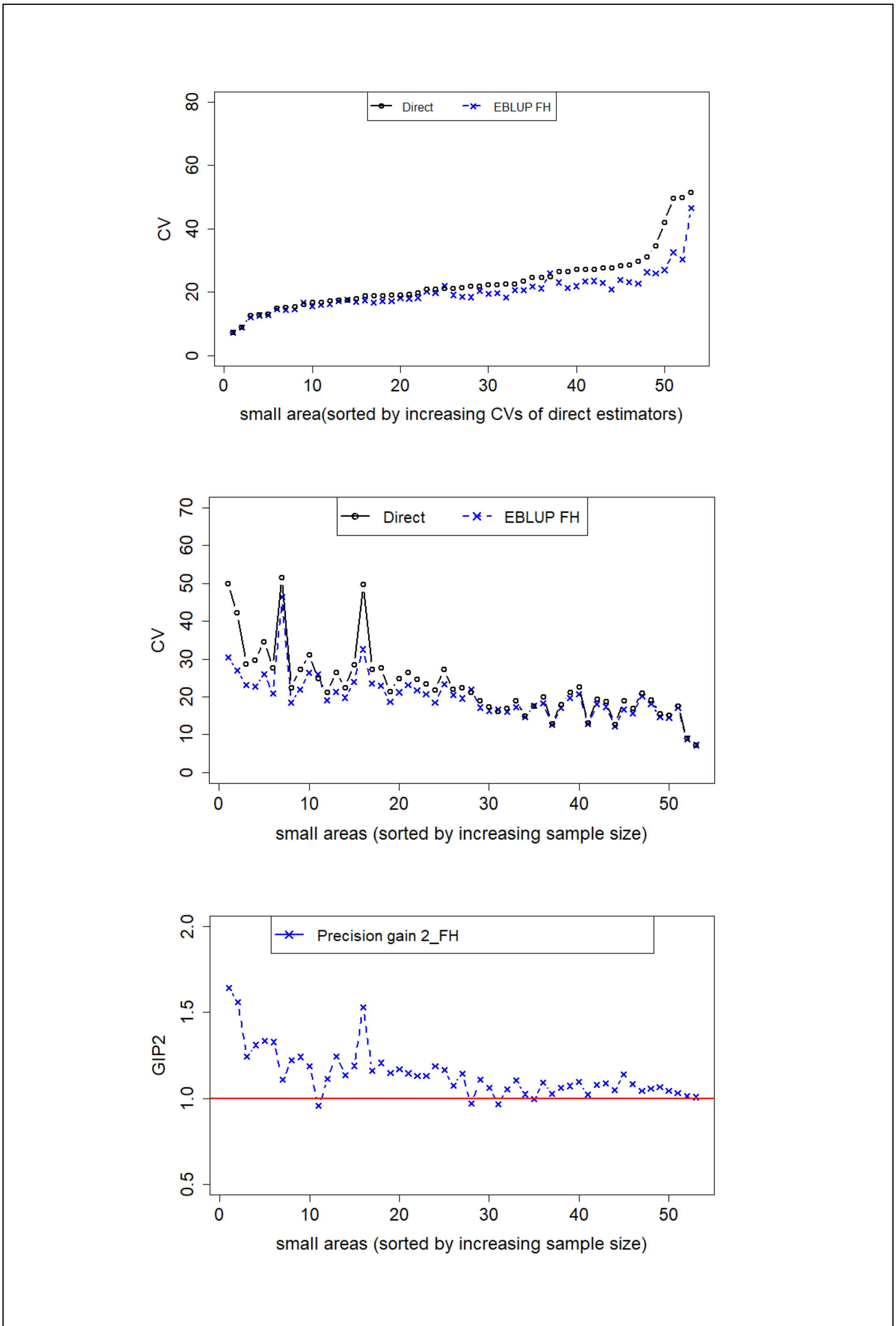

Figure 4.7.31 Coefficients of variation for the Direct and Fay-Herriot estimator of the poverty gap in Greece for the year 2009 (sorted by increasing sample size and sorted by increasing CVs of direct estimator) and gain in precision index (GIP2)

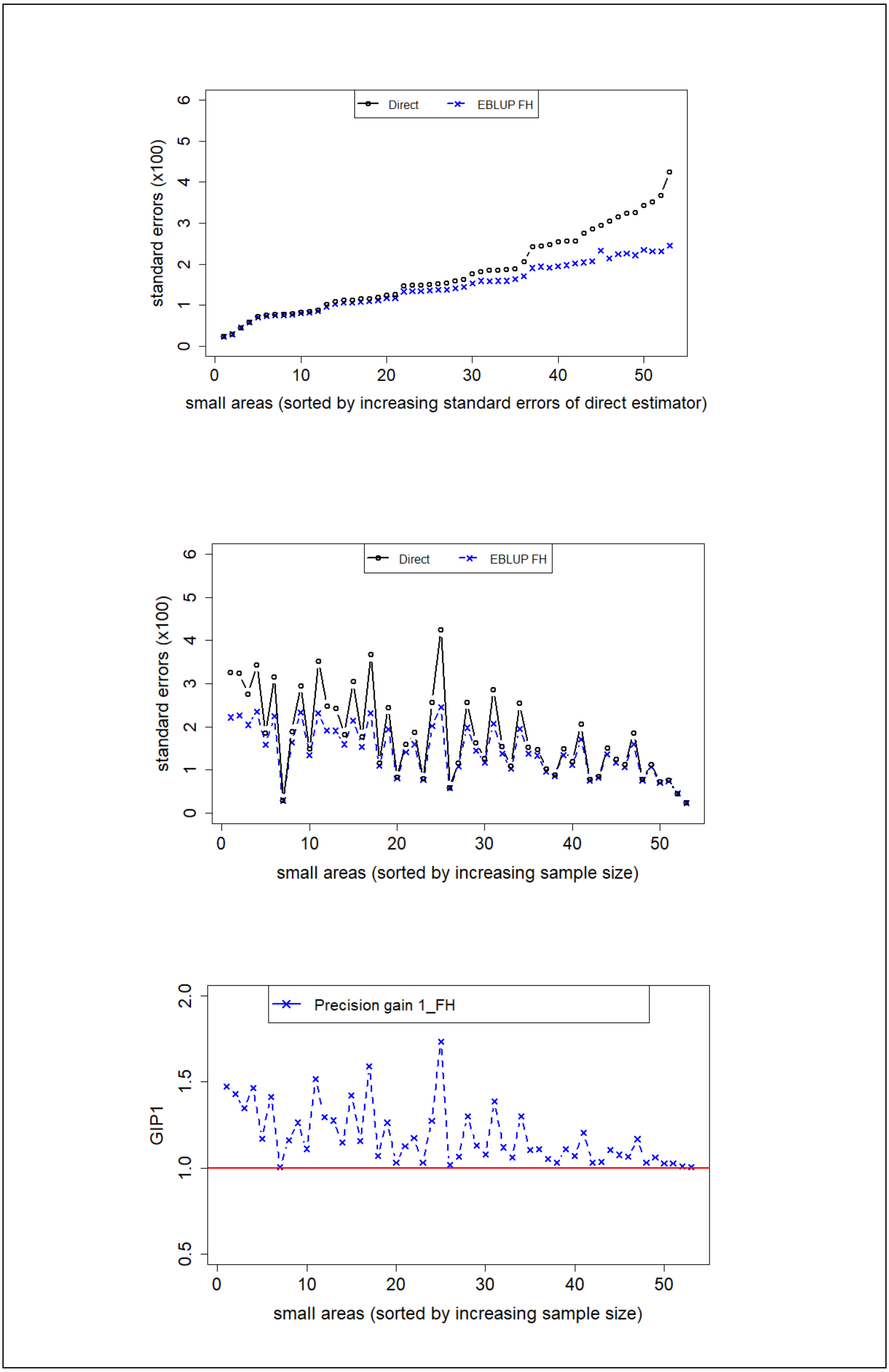

Figure 4.7.32 Standard errors for the Direct and Fay-Herriot estimator of the poverty gap in Greece for the year 2009 and gain in precision index (GIP1) sorted by increasing sample size

4.8 Results

4.8.1 Results for the estimation of headcount ratio and poverty gap in Greece for the year 2013. The EBLUP Fay-Herriot estimates of the headcount ratio and the poverty gap in Greece for the year 2013 are given in detail in Tables A7, A8 in the Appendix. Also, direct and Fay-Herriot estimates of the headcount ratio and the poverty gap were placed on the map (Figures 4.8.1 - 4.8.4) to illustrate the spatial distribution of the analyzed phenomena. Furthermore Figures 4.8.5 and 4.8.6 show the direct and Fay-Herriot estimates of the headcount ratio and the poverty gap respectively in relation to the area specific sample size as well as in the relation to the coefficient of variation (CV) of the direct estimator.

Concerning the headcount ratio:

- The highest poverty rates occur in the prefectures of Imathia (39.01%), Xanthi (38.16%), Etolia and Akarnania (35.95%, Rodopi (32.8%), West Attiki (32,7%) and Pella (31,88%).
- The lowest poverty rates occur in the prefectures of Fokida (6.7%), Viotia (14.56%), Pireas (14.58%) and Samos (17.76%).
- The biggest differences in relation to direct estimates are observed in the prefectures of Thesprotia (30,14%), Samos (26.61%), Kefalinia (11,74%), Lasithi (7,36%), Imathia (7,32%) and Grevena (6.52%). In these areas (except the prefecture of Imathia) the sample size is small (n< 70).
- In 10 prefectures the difference between direct and Fay-Herriot estimates is over 3%, while in 16 prefectures the difference is over 2%. In these areas there is a reduction (in most areas quite large) for both the MSE and the CV of estimates.

Concerning the poverty gap, which measures the degree of poverty for people under the poverty line, the conclusions in most cases are the same:

- The highest poverty gaps rates occur in the prefectures of Imathia (17.68%), Xanthi (16.79%), Etolia and Akarnania (13.38%), Rethymno (12.51%), West Attiki (12.47%) and Pieria (11.99%). The lowest poverty gaps rates occur in the prefectures of Fokida (0.75%), Viotia (1.79%), Lefkada (2.58%) and Lakonia (3.02%).
- The biggest differences in relation to direct estimates are observed in the prefectures of Samos (19.65%), Thesprotia (8.7%), Lassithi (7.40%), Imathia

(6.7%) and Kefalinia (4.84%). In these areas (except the prefecture of Imathia) the sample size is small (n< 70).

- It is worth mentioning the fact that in the prefectures of Lefkada and Lakonia while the poverty rate is close to 20% the poverty gap rate is very low close to 3%. For instance, in Lefkada the poverty gap is estimated at 2.56% of the at-risk-of-poverty threshold and this means that 50% of the poor population in this prefecture has an income higher than 97.44% of the at risk of poverty threshold (5,023 euros), that is to say more than 4,894.41 euros, yearly, per person. Also, in the prefecture of Imathia the poverty gap is estimated at 17.68% of the at-risk-of-poverty threshold and this means that 50% of the poor population in this prefecture has an income higher than 82.32% of the at risk of poverty threshold (5,023 euros), that is to say more than 4,134.93 euros, yearly, per person.

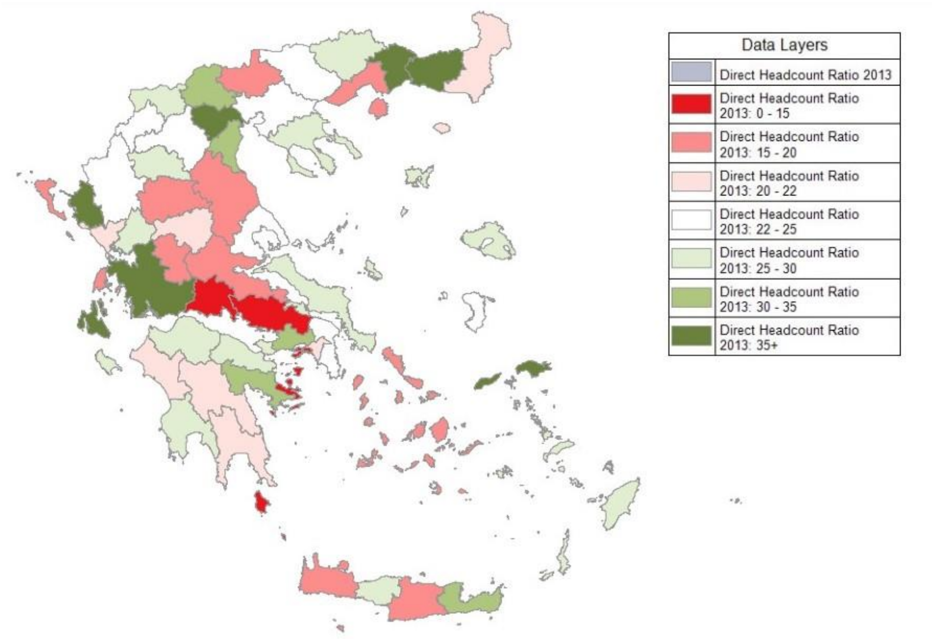

Figure 4.8.1 Cartogram of direct estimates of the headcount ratio in Greece for the year 2013

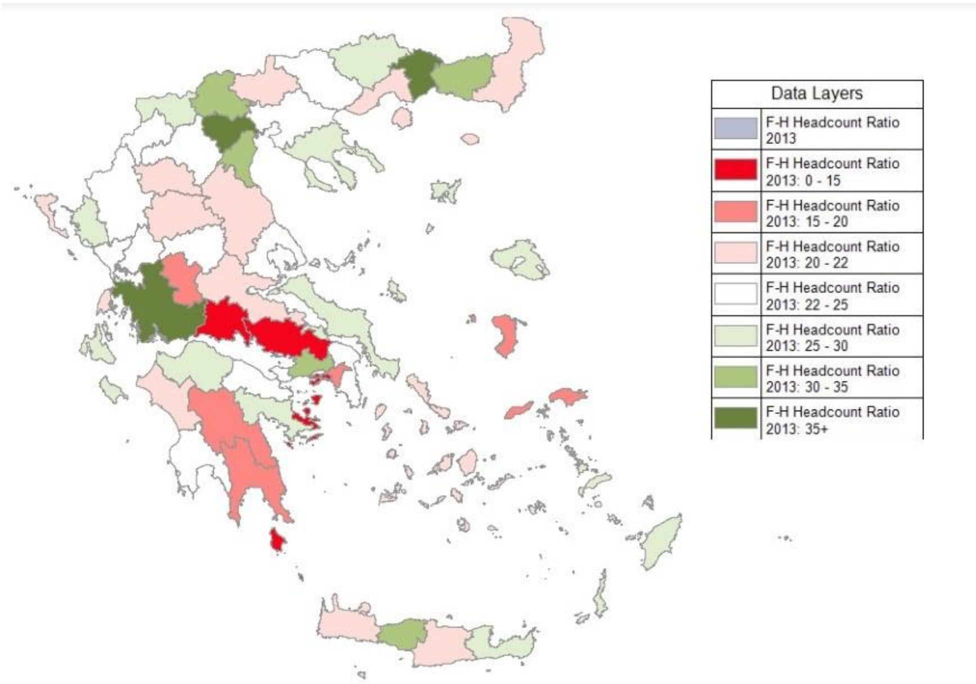

Figure 4.8.2 Cartogram of EBLUP Fay-Herriot estimates of the headcount ratio in Greece for the year 2013

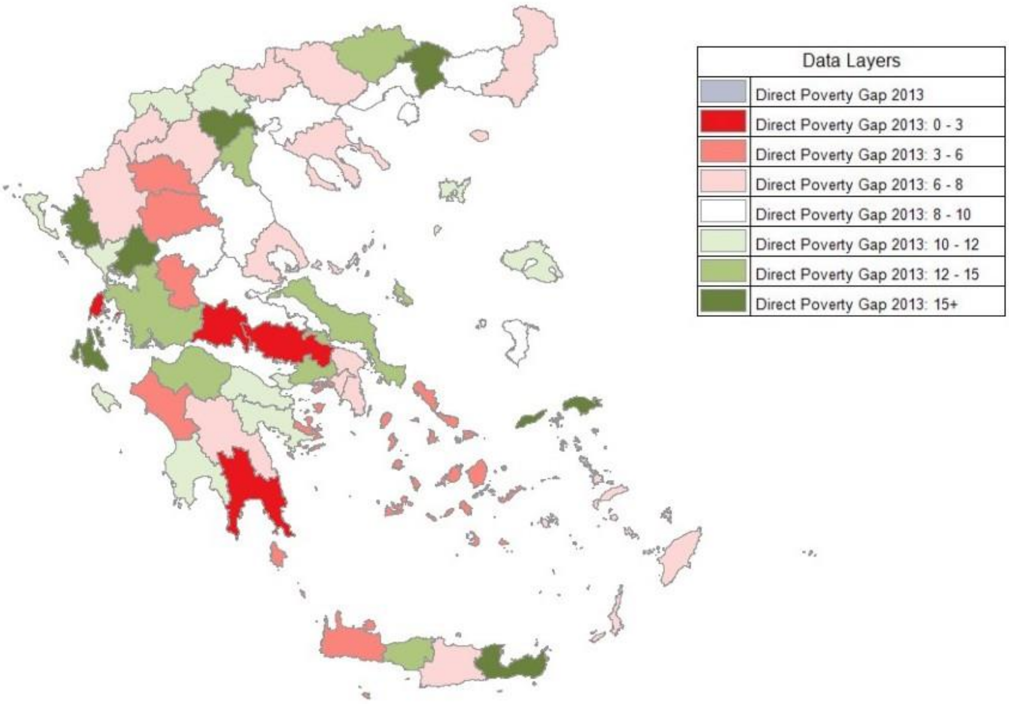

Figure 4.8.3 Cartogram of direct estimates of the poverty gap in Greece for the year 2013

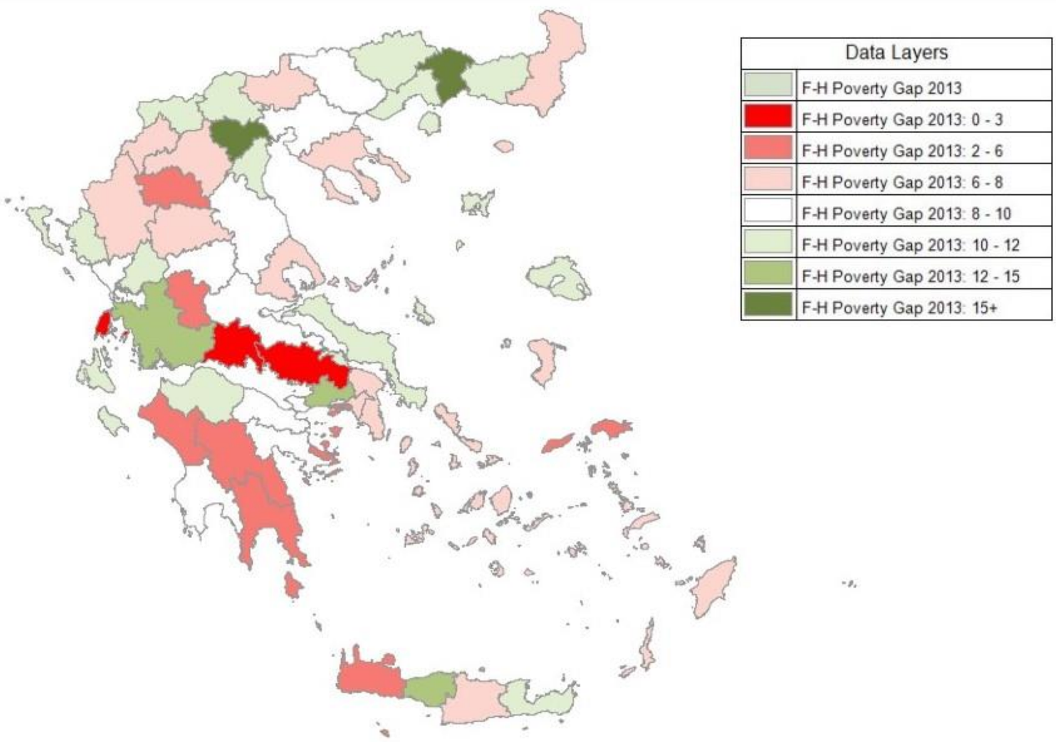

Figure 4.8.4 Cartogram of EBLUP Fay-Herriot estimates of the poverty gap in Greece for the year 2013

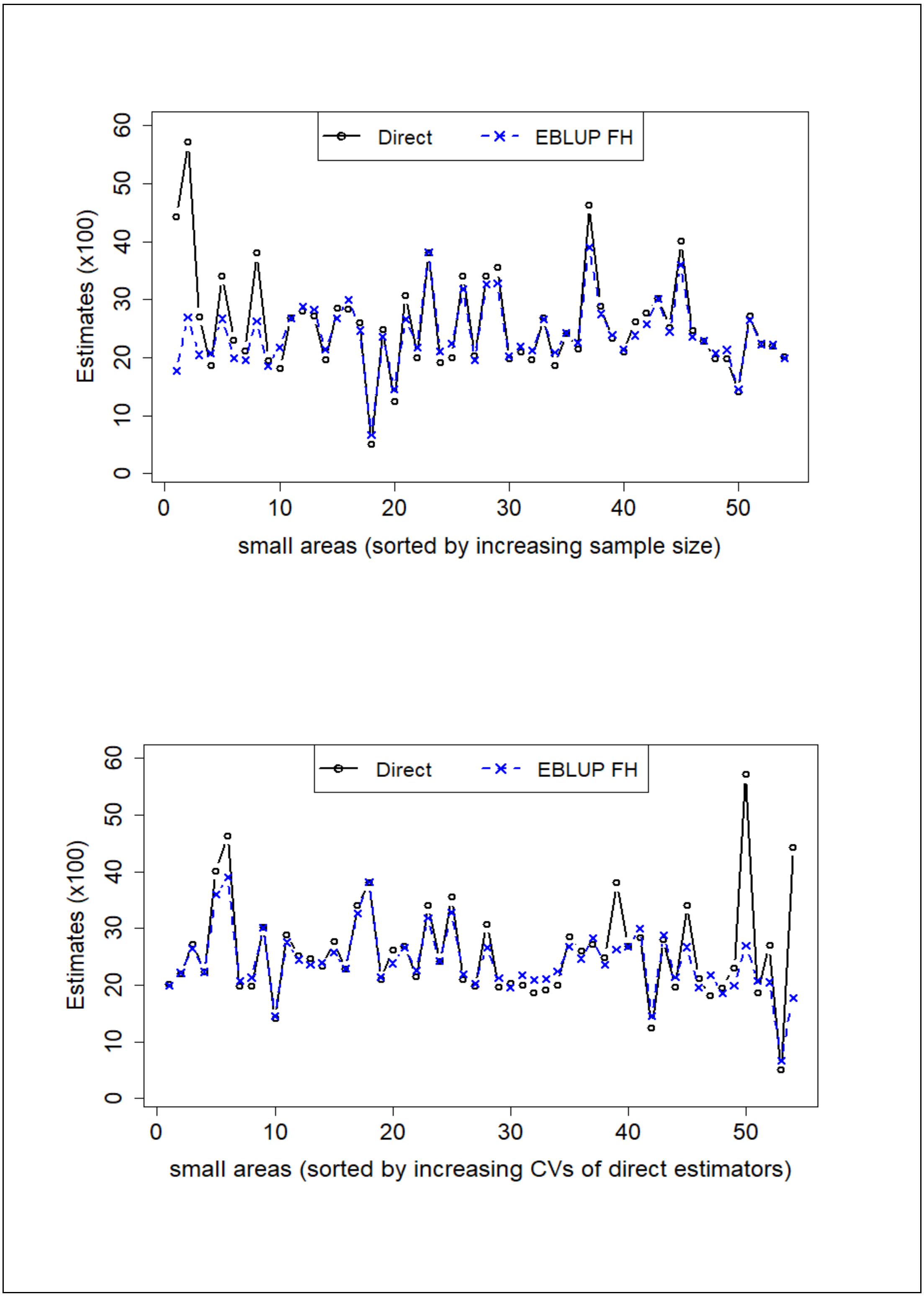

Figure 4.8.5 Direct and Fay-Herriot estimates of the headcount ratio in Greece for the year 2013 sorted by increasing sampling size and by CVs of direct estimators

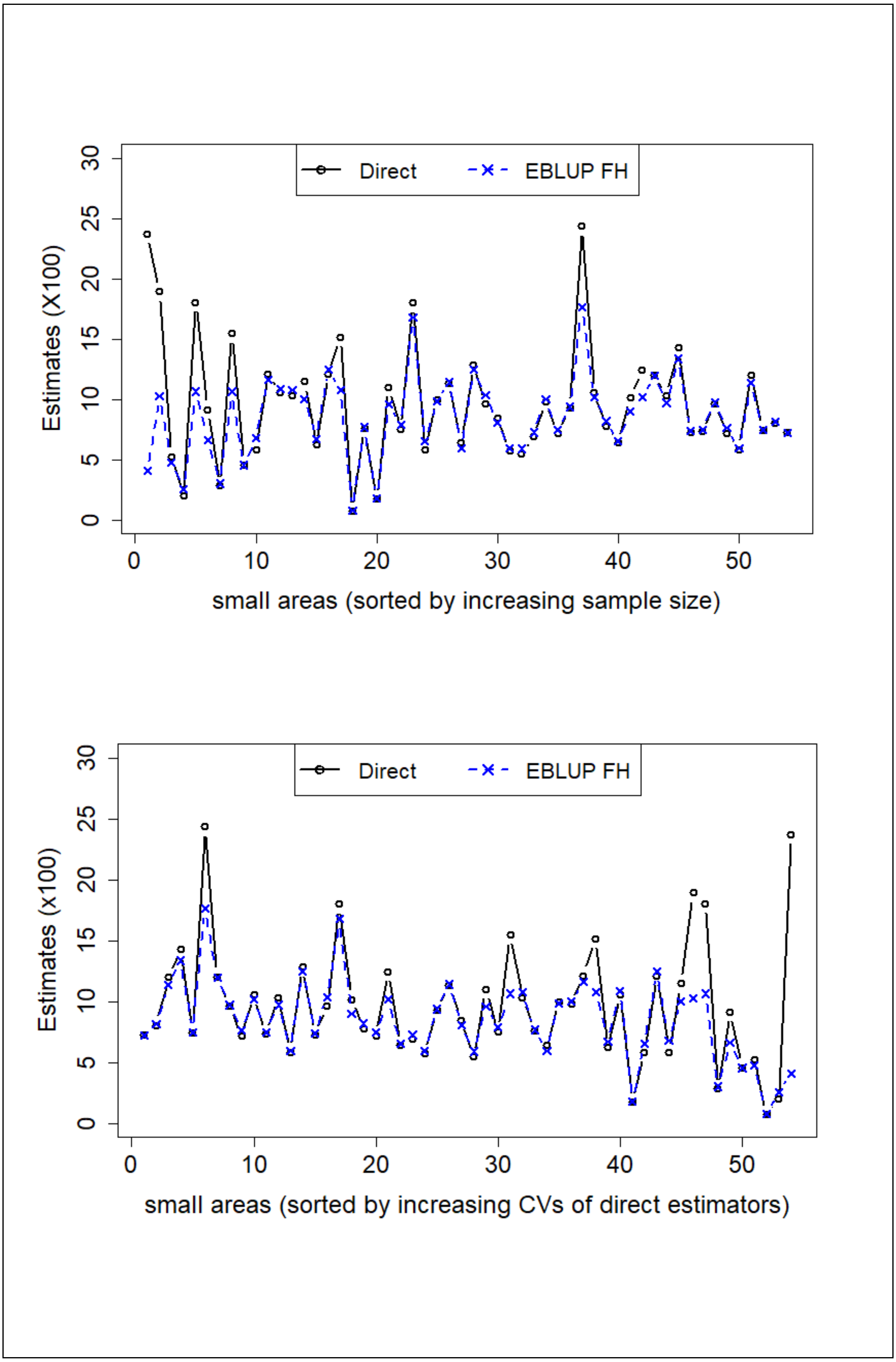

Figure 4.8.6 Direct and Fay-Herriot estimates of the poverty gap in Greece for the year 2013 sorted by increasing sampling size and by CVs of direct estimators

4.8.2 Results for the estimation of headcount ratio and poverty gap in Greece for the year 2009. The EBLUP Fay-Herriot estimates of the headcount ratio and the poverty gap in Greece for the year 2009 are given in detail in Tables A9, A10 in the Appendix. Also, direct and Fay-Herriot estimates of the headcount ratio and the poverty gap were placed on the map (Figures 4.8.7 - 4.8.10) to illustrate the spatial distribution of the analyzed phenomena. Furthermore Figures 4.8.11 and 4.8.12 show the direct and Fay-Herriot estimates of the headcount ratio and the poverty gap respectively in relation to the area-specific sample size as well as in relation to the coefficient of variation (CV) of direct estimator. According to the EU-SILC survey in Greece for the year 2013 the headcount ratio of the total of population was 19.7% and the poverty gap was 24.1%.

Concerning the headcount ratio:

- The highest poverty rates occur in the prefectures of Serres (43.11%), Kilkis (39.72%), Lesvos (37.95%), Grevena (36.66%), Xanthi (36.03%) and Evia (35,34%). The lowest poverty rates occur in the prefectures of Rethymno (4.87%), Lassithi (10.41%), East Attiki (10.98 %) and Kyklades (11.87%).
- The biggest differences in relation to direct estimates are observed in the prefectures of Zakyntho (19.03%), Messinia (13.73%), Evrytania (10.86%), Xanthi (10.55%) and Samos (8.49%). From these areas, the prefectures of Zakynthos, Evrytania and Samos have a small sample size (n<36).
- In 15 prefectures the difference between direct and Fay-Herriot estimates is over 3%. In these areas there is a reduction (in most areas quite large) for both the MSE and the CV of estimates.

Concerning the poverty gap, which measures the degree of poverty for people under the poverty line, the conclusions in most cases are the same:

- The highest poverty gaps rates occur in the prefectures of Serres (13.4%), Kilkis (13.3%), Evia (12.36%), Ilia (11.2%), Drama (10.94%) and Grevena (10.7%). The lowest poverty gap rates occur in the prefectures of Rethymno (0.62%), Kyklades (2.81%), Athens (3.18%) and Ioannina (3.67%).
- The biggest differences in relation to direct estimates are observed in the prefectures of Evia (5.39%), Xanthi (5.24%), Messinia (5.09%) and Florina (3.62%).

- The prefectures of Serres, Kilkis and Evia with some of the highest poverty rates also have some of the highest poverty gap rates. Respectively, the prefectures of Rethymno and Kyklades with some of the lowest poverty rates also have some of the lowest poverty gap rates. For instance, in the prefecture of Serres the poverty rate is 43.11% and the poverty gap rate is 13.4%. This mean that 50% of the poor population in this prefecture has an income higher than 86,6% of the at risk of poverty threshold (6,897 euros), that is to say more than 5972.802 euros, yearly, per person. Also, in the prefecture of Rethymno the poverty rate is 4.87% and the poverty gap rate is 0.62%. This mean that 50% of the poor population in this prefecture has an income higher than 99.38% of the at risk of poverty threshold (6,897 euros), that is to say more than 6,854.2 euros, yearly, per person.

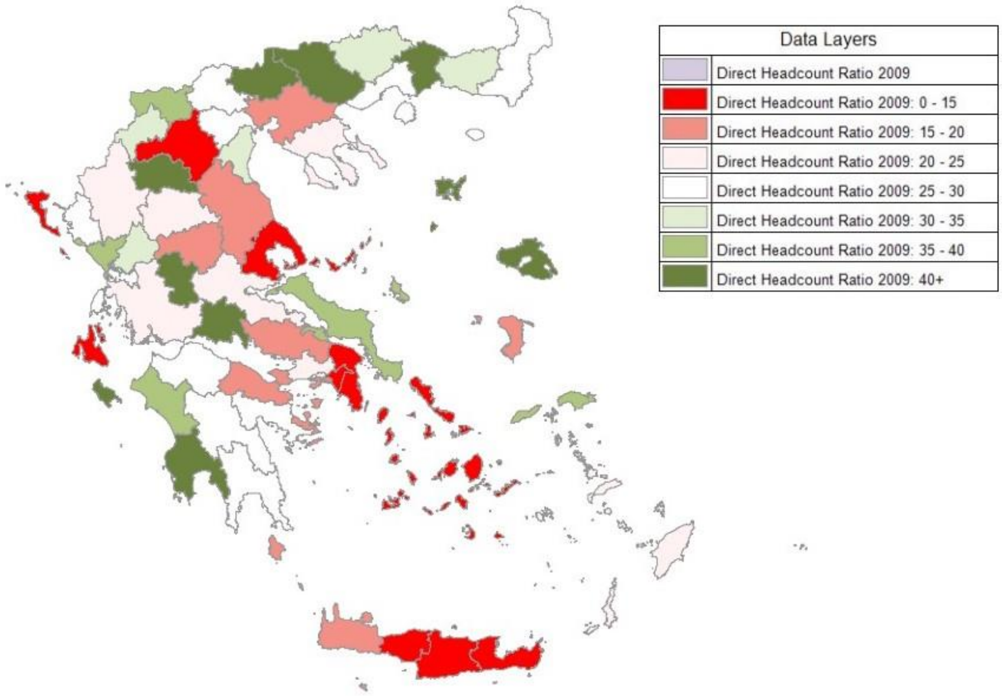

Figure 4.8.7 Cartogram of direct estimates of the headcount ratio in Greece for the year 2009

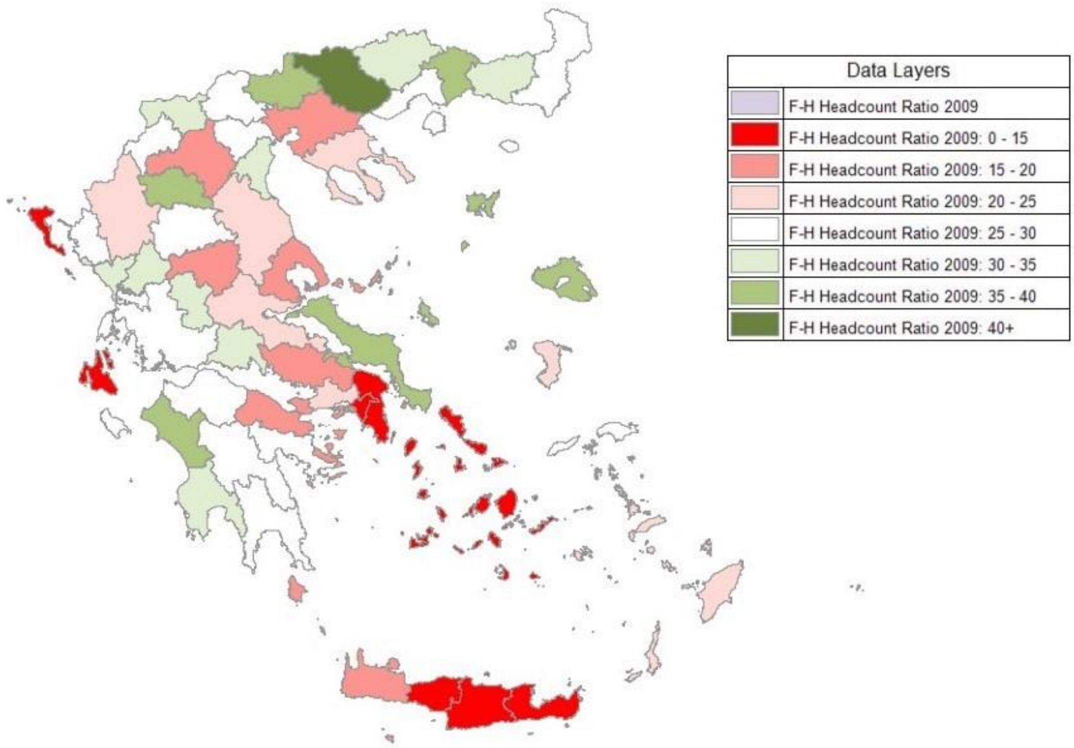

Figure 4.8.8 Cartogram of EBLUP Fay-Herriot estimates of the headcount ratio in Greece for the year 2009

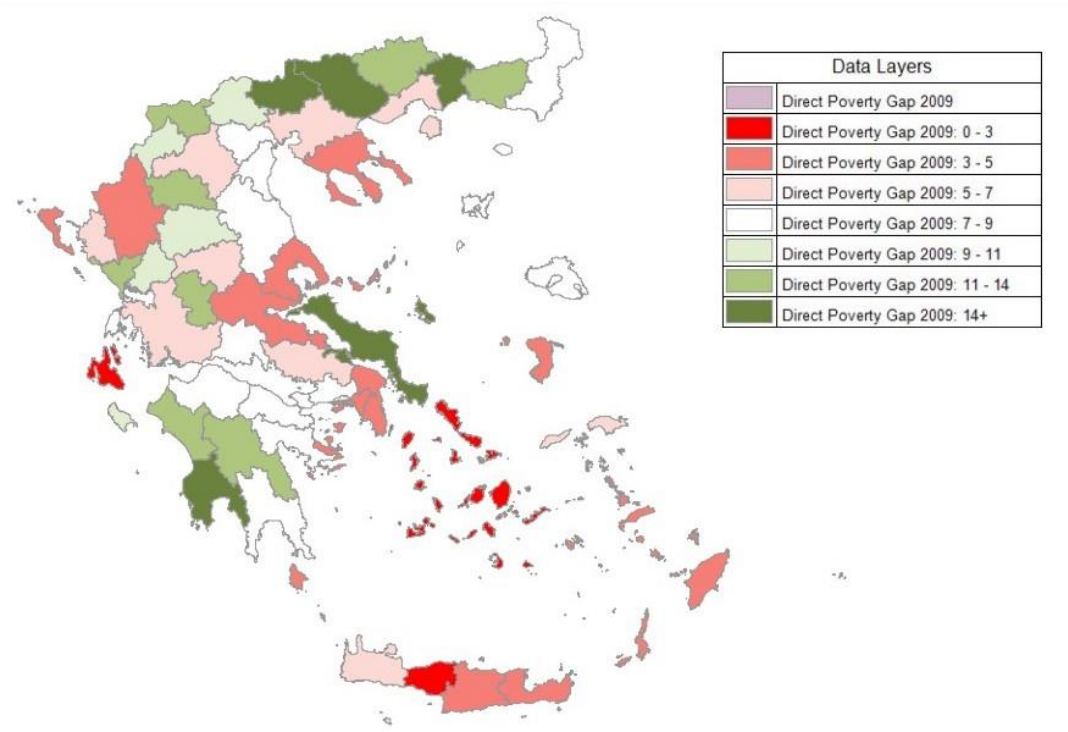

Figure 4.8.9 Cartogram of direct estimates of the poverty gap in Greece for the year 2009

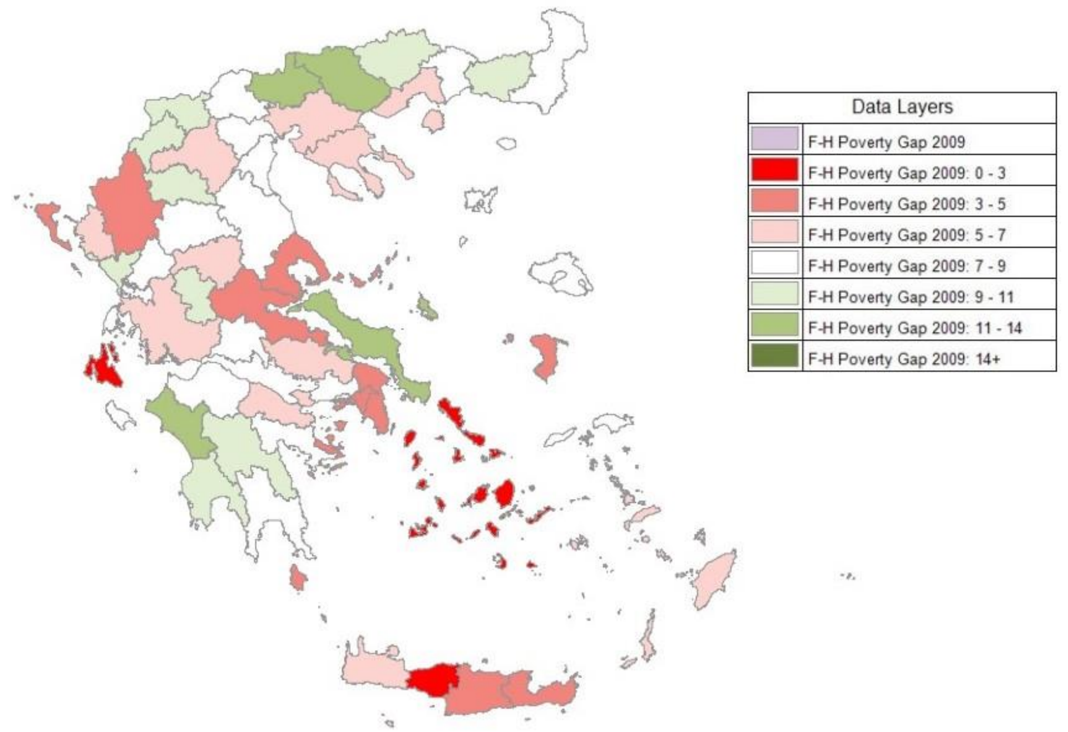

Figure 4.8.10 Cartogram of EBLUP Fay-Herriot estimates of the poverty gap in Greece for the year 2009

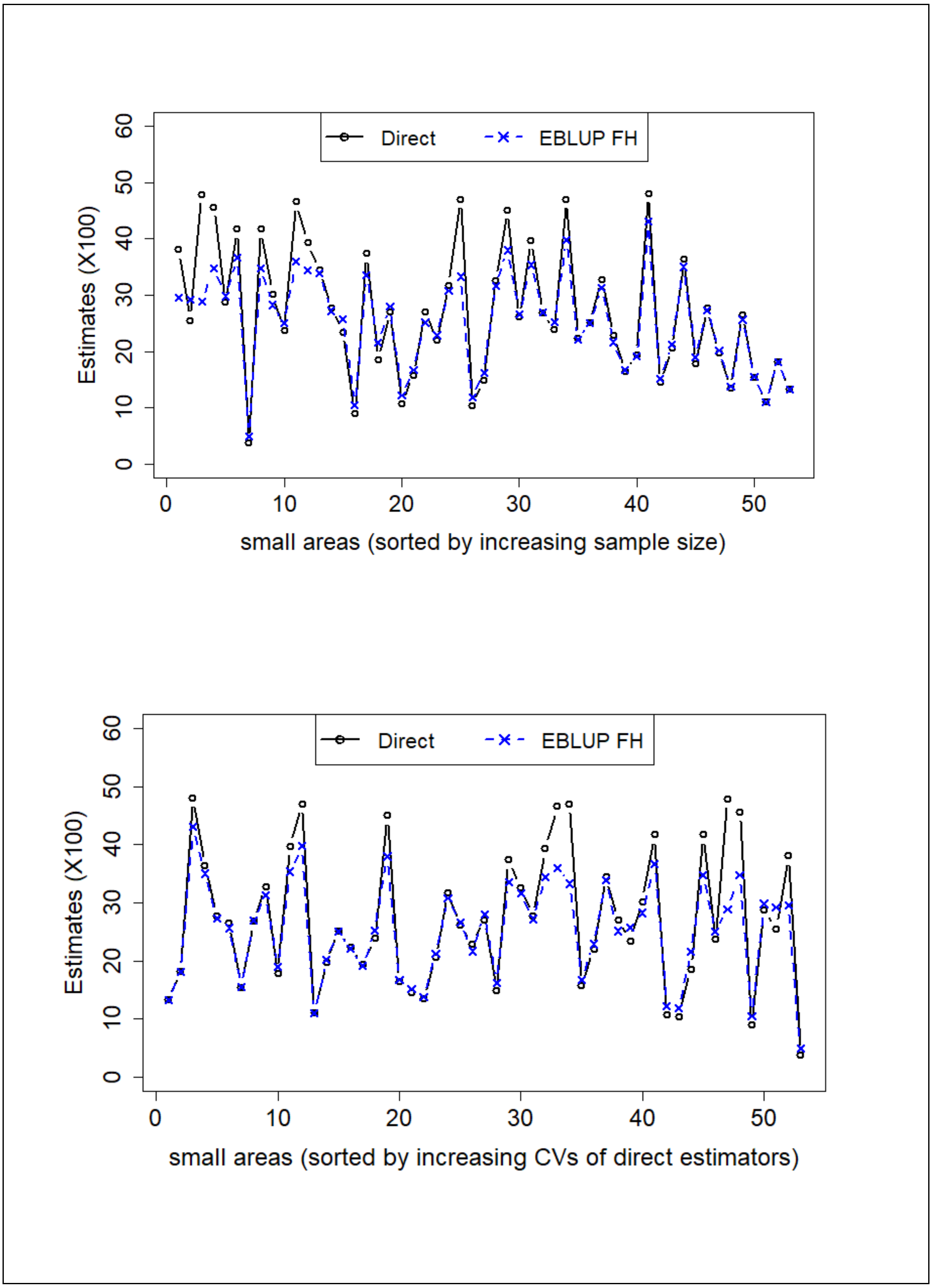

Figure 4.8.11 Direct and Fay-Herriot estimates of the headcount ratio in Greece for the year 2009 sorted by increasing sampling size and by CVs of direct estimators

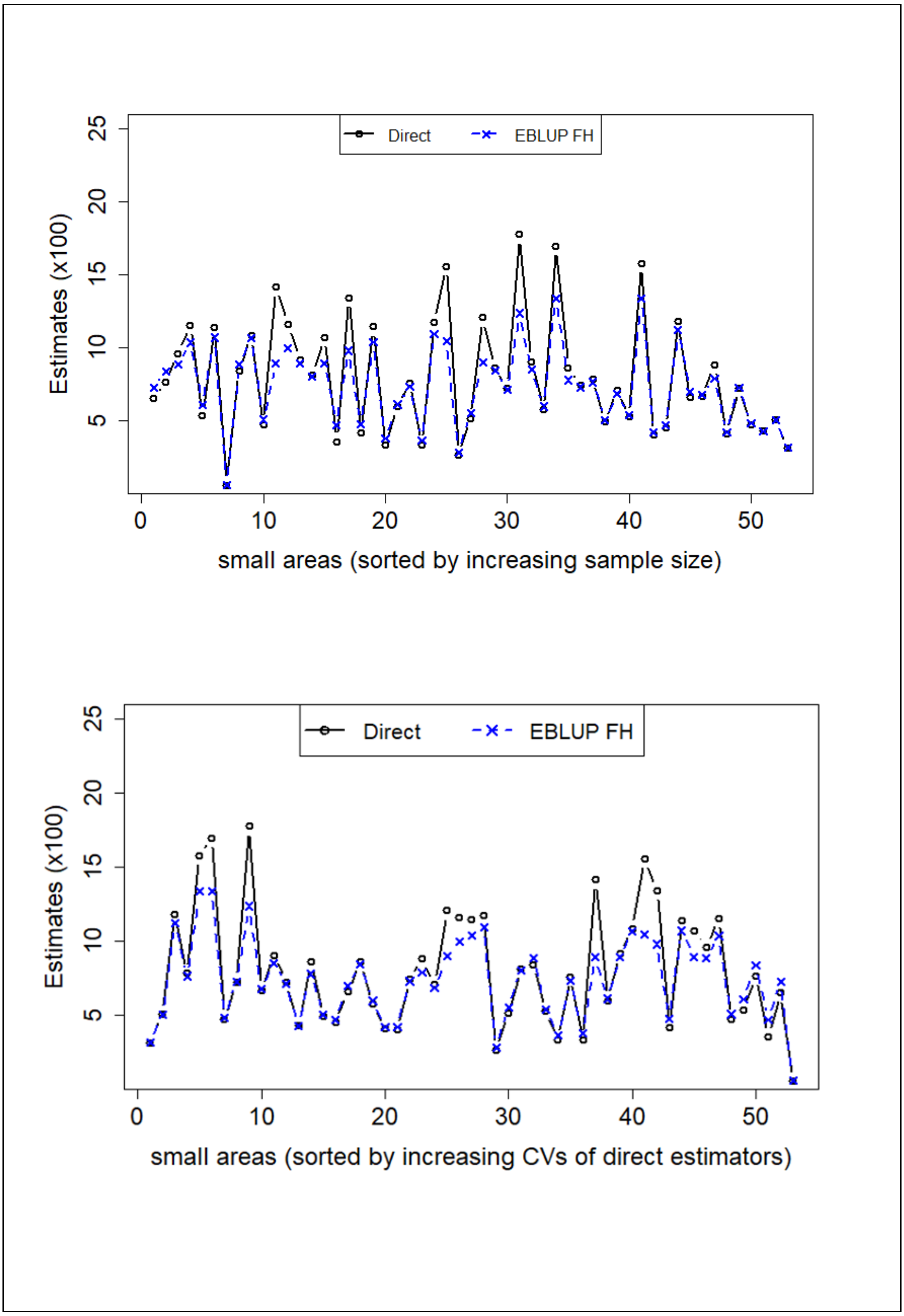

Figure 4.8.12 Direct and Fay-Herriot estimates of the poverty gap in Greece for the year 2009 sorted by increasing sampling size and by CVs of direct estimators

4.8.3 Comparative results for the estimation of poverty in Greece for the years 2009 and 2013.

The EBLUP Fay-Herriot estimates of the headcount ratio in Greece for the years 2009 and 2013 are given in detail in Table A21 in the Appendix. Also, the EBLUP Fay-Herriot estimates of the headcount ratio and the poverty gap for the year 2013 compared to the year 2009 are presented in Figures 4.8.13 - 4.8.16. Furthermore, a scatter plot was created for the data of 2009 versus 2013, both for the headcount ratio and for the poverty gap, to examine whether there is a pattern between the two years of data (Figures 4.8.17 and 4.8.18).

Concerning the headcount ratio:

- The biggest absolute differences occurred in the prefectures of Fokida (28.05%), Rethymno (25,13%), Serres (19.21%) and Kilkis (17.9%).
- The smallest absolute differences occurred in the prefectures of Zakynthos (0.16%), Larissa (0.69%), Ioannina (0.73%), Achaia (0.78%) and Prefecture of Pireas (0.94%).
- In 16 prefectures the difference between 2009 and 2013 was over 10%. Also, in 30 prefectures there was a reduction in the poverty rate for the year 2013 compared to 2009.
- Based on the scatter plot (Figure 4.8.17) there is no trend to the data.

As far as the poverty gap is concerned:

- The biggest absolute differences occurred in the prefectures of Rethymno (11.88%), Imathia (10.37%), Fokida (8.1%) and Xanthi (7.8%).
- The smallest absolute differences occurred in the prefectures of Preveza (0.1%), Chania (0.54%), Evros (0.61%) and Drama (0.69%).
- In 22 prefectures the difference between 2009 and 2013 was over 4%. Also, in 36 prefectures there was an increase in the poverty gap rate in the year 2013 compared to 2009.
- Based on the scatter plot (Figure 4.8.18) there is no trend to the data.

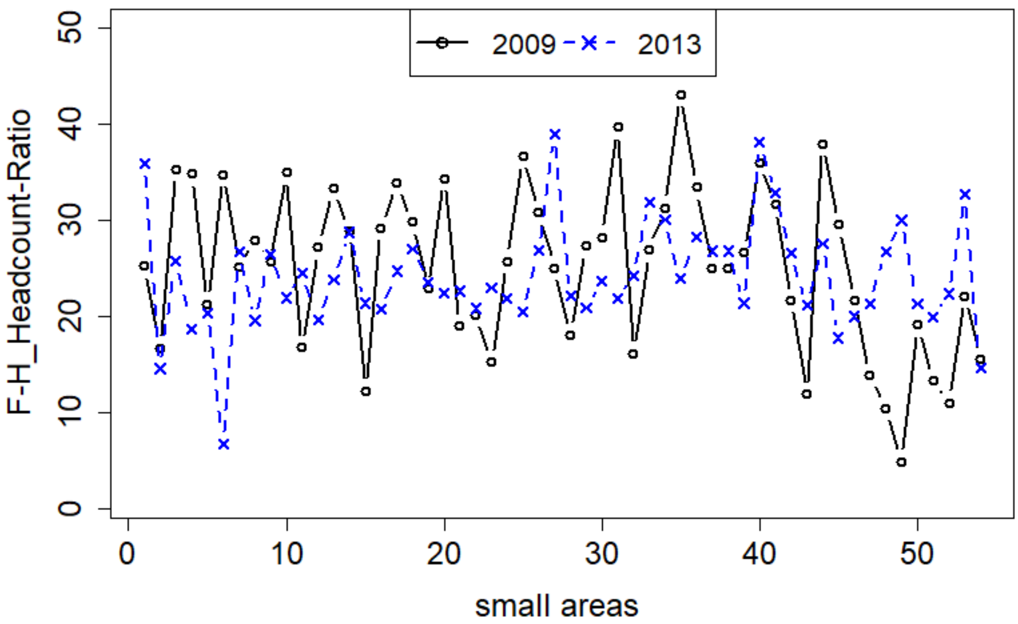

Figure 4.8.13 EBLUP F-H estimates of the headcount ratio in Greece for the years 2009 and 2013

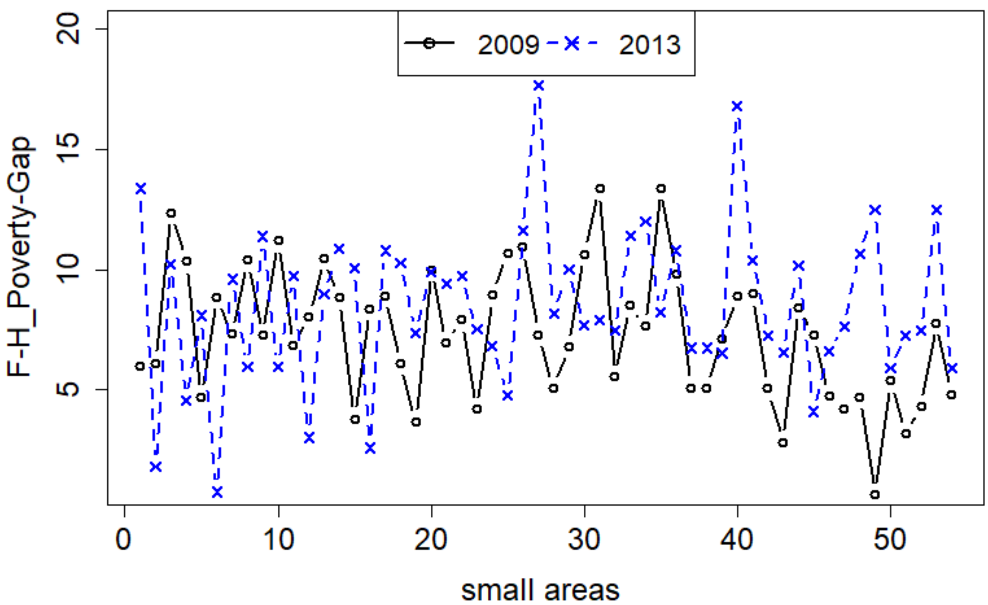

Figure 4.8.14 EBLUP F-H estimates of the poverty gap in Greece for the years 2009 and 2013

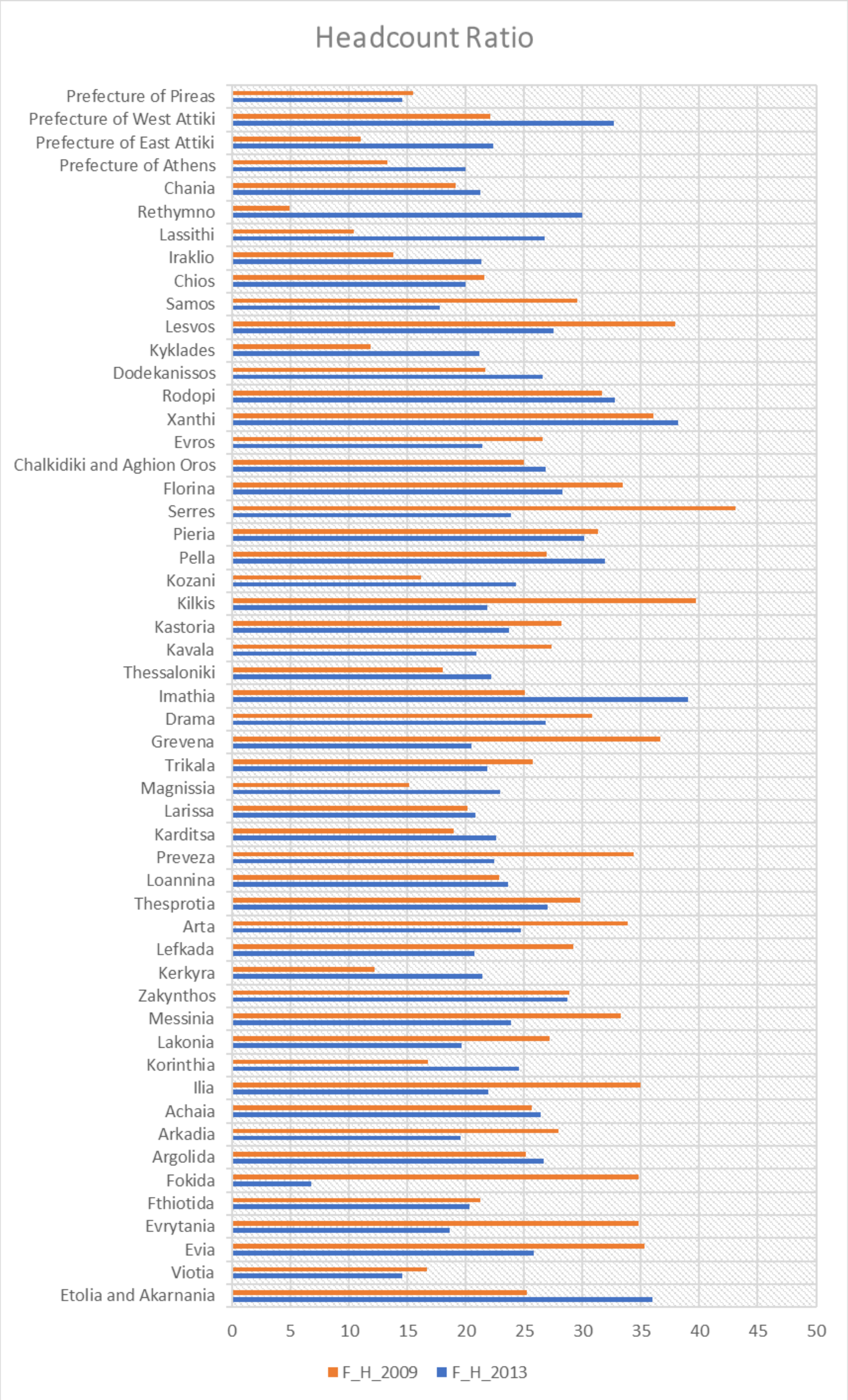

Figure 4.8.15 Comparative bar chart for the EBLUP F-H estimates (%) of the headcount ratio in Greece for the years 2009 and 2013

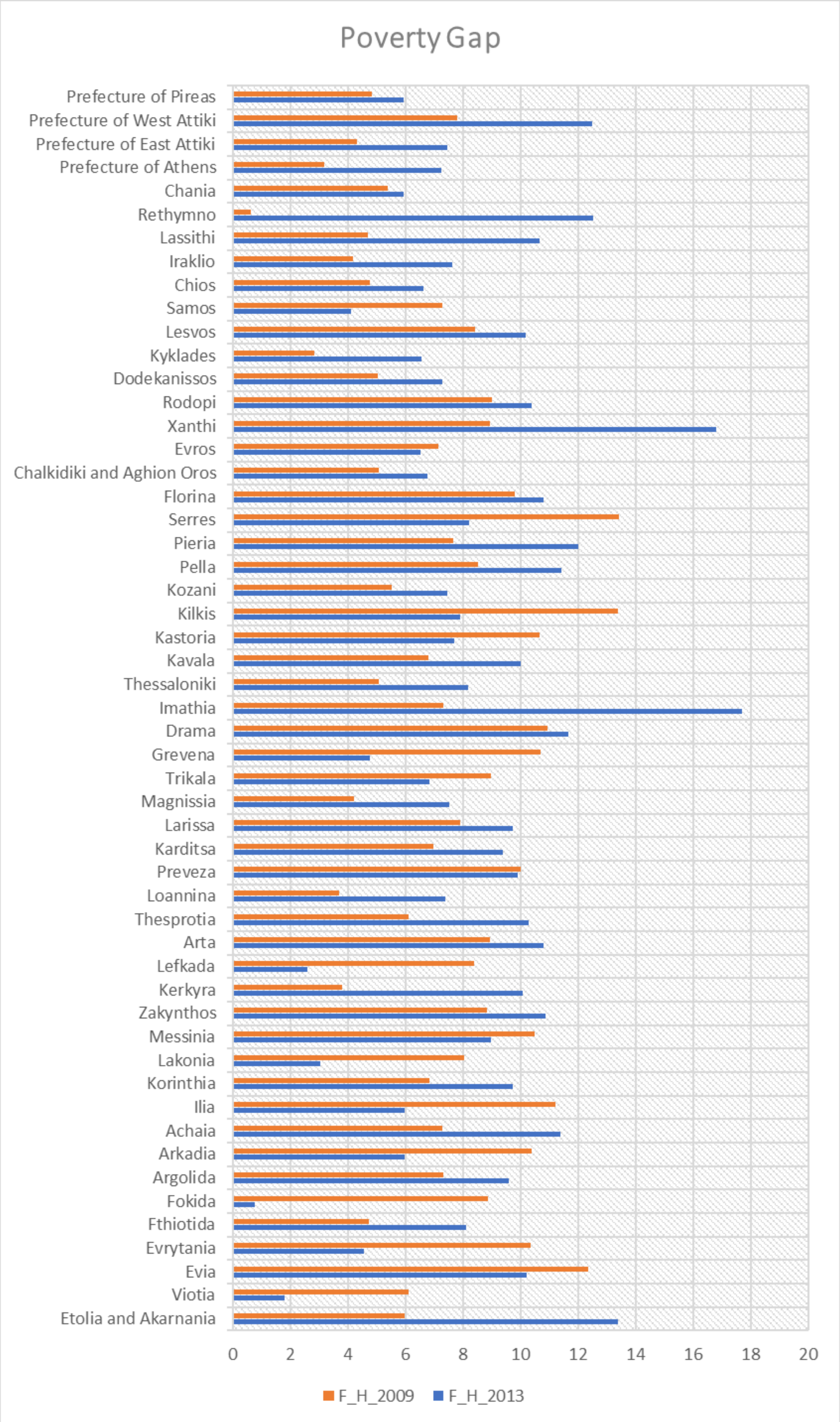

Figure 4.8.16 Comparative bar chart for the EBLUP F-H estimates (%) of the poverty gap in Greece for the years 2009 and 2013

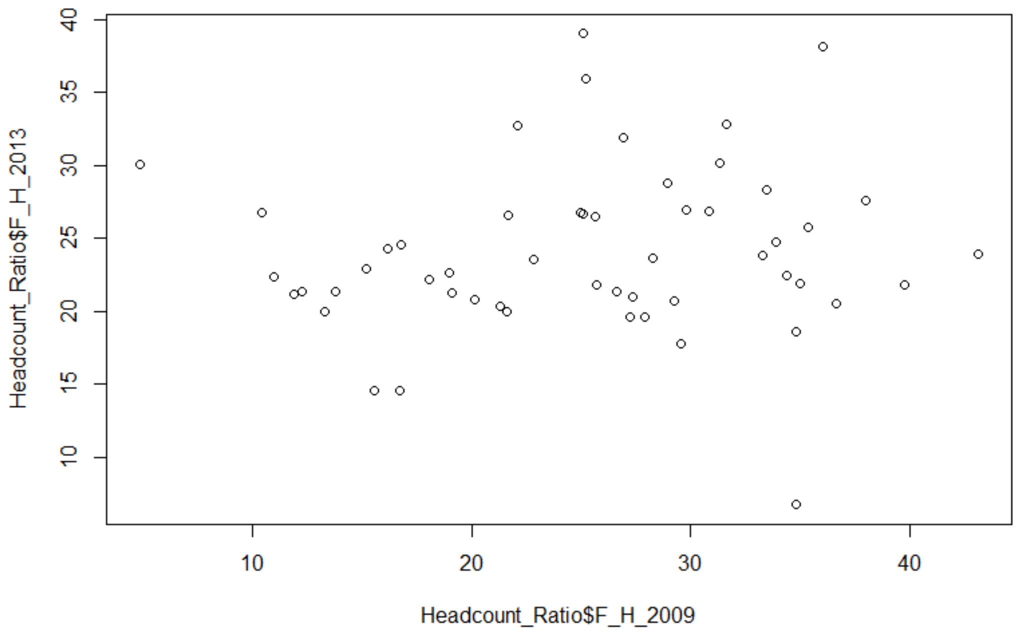

Figure 4.8.17 Scatter plot for the headcount ratio in 2009 versus the headcount ratio in 2013

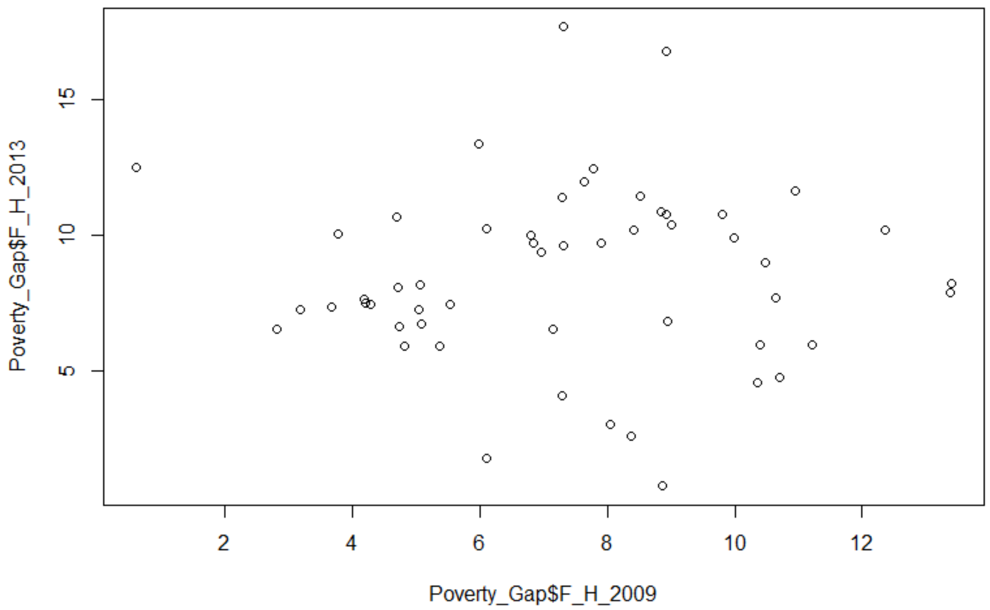

Figure 4.8.18 Scatter plot for the poverty gap in 2009 versus the poverty gap in 2013

5. Estimating Unemployment Using Small Area Estimation Methods

5.1 Introduction

Labour market statistics are crucial for setting policy objectives, choosing between strategic choices, assisting in policy making and monitoring policy effectiveness (Sengenberger, 2011). The structure of the economically active population and, in particular, unemployment and its structure have always been of great social interest. Of all the labor indicators, unemployment is by far the most frequently reported in the media in almost every country in the world (Hussmanns, 2007). Specially in recent decades when unemployment has assumed alarming dimensions both in Greece and globally. The fight against unemployment is one of the goals of the European Commission, which in the framework of the "Europe 2020" strategy set a target of 75% employment rate for the EU population aged 20-64 (European Commission, 2010).

Statistics on the economically active population and its components can be produced from a variety of sources. Common sources of statistics for the economically active, employed and unemployed are labor force surveys and other household sampling surveys, as well as population Censuses. However, in most of these surveys the sample size allows an accurate estimate of unemployment only at a very general level such as the whole country and regions (NUTS 2). As unemployment is characterized by large territorial differences at both national and regional level, its estimation at a lower spatial level seems crucial. Therefore, small area estimation techniques for measuring unemployment at local level are required. Under the Fifth Framework Program of the European Union, Eurostat funded a research project (EURAREA, 2004) which provided information on how SAE methods are used in the production of official statistics. Unemployment was one of the target variables examined in each country participating in the program.

5.2 Unemployment measurement

Two critical issues concern the measurement of unemployment. One is related to the problem of identifying unemployment and the other to the problem of constructing an appropriate index of overall unemployment using the information available on the unemployed (Paul, 1991).

5.2.1 Definitions of unemployment. Statistical measurement of unemployment should start with a clear definition. Once the phenomenon of unemployment has been defined a target population can be identified and the relevant parameters of the population can be determined (Bartholomew, Moore, Smith and Allin, 1995). In earlier times when it was the norm for there to be a principal breadwinner in each household the concept of unemployment was definitely simpler, as this person usually worked full time. In those days the number of unemployed corresponded, at least approximately, to social reality. Nowadays the situation is more fluid, and it is no longer possible to draw a simple distinction between the employed and the unemployed. The employment situation is a complex process in which people move in and out of different "states" over time. Therefore, a complete description should reflect this complexity.

In the literature, several criteria have been discussed for identifying unemployment. Paul (1991) distinguishes the following:

- Minimum level of income criteria
- Criterion of productivity
- Recognition criterion and
- Time criterion

According to the income criterion (Dandekar and Rath, 1971) a person can be considered unemployed if their income is less than a "minimum level". The minimum income level can either be set arbitrarily or correspond to the socially acceptable poverty line. A second approach to determining unemployment is based on the productivity criterion. According to this criterion, a person can be considered unemployed if their marginal productivity is lower than a certain cut-off level. The cut-off level is often taken at zero. Furthermore, Sen (1975) proposed a criterion based on what he defines as the "recognition criterion". In this approach, an employed person may identify themselves as unemployed if their employment does not meet their expectations regarding self-esteem or the full use of their training. That is, a person may be considered unemployed if they are not satisfied with their work. In addition, according to the time criterion a person can be considered unemployed if their actual days of employment are less than the actual working days during the reference period.

The current international standards for labour force statistics include the Resolution concerning statistics of the economically active population, employment, unemployment and underemployment adopted by the International Labour Organization (ILO)[70] and specifically by the Thirteenth International Conference of Labour Statisticians (ICLS)[71] in October 1982. The definitions of unemployment adopted were as follows (ILO, 2013):

- In Employment: All aged 15[72] and over who did some paid work in the reference week (whether as employed or self-employed); those who had a job from which they were temporarily absent and those on government employment and training schemes. Also, those who work in a family business without pay are recognized as a separate group in employment. The concept of "some work" is interpreted as work for at least one hour during the reference period.
- Unemployed: A person who has no work for pay or kind in the reference week and who is able to start work in the next two weeks and who has been actively seeking work during the last four weeks or is waiting to start a job already obtained.
- Economically inactive: Those not in employment nor unemployed in the ILO sense including those under 15, those looking after family, or those retired.

Also, the labour force (formerly known as the economically active population) is the sum of the number of persons employed and the number of persons unemployed.

5.2.2 Indicators of unemployment. Once a definition of unemployment has been formulated, the next step is to construct an appropriate index of overall

---

[70] The International Labour Organisation (ILO) publishes international standards on the various topics of labour statistics. These standards are set by the International Conference of Labour Statisticians (ICLS), which is convened by the ILO about every five years. They include the ILO Labour Statistics Convention, 1985 (No. 160), the ILO Labour Statistics Recommendation, 1985 (No. 170), and the various Resolutions adopted by the ICLS on specific topics of labour statistics. The purpose of the ICLS Resolutions is to provide technical guidelines for the development of national labour statistics on the basis of accepted definitions and methods, to enhance the international comparability of labour statistics, and to protect labour statistics against public criticism and political interference at the national level. Resolution concerning statistics of work, employment and labour underutilization, adopted by the 19th International Conference of Labour Statisticians, Geneva, October 2013; http://www.ilo.org/global/statistics-and- databases/standardsand-guidelines/resolutions-adopted-by-international-conferences-of-labour-statisticians/WCMS_230304/lang-- en/index.htm.

[71] The Resolution concerning the measurement of underemployment and inadequate employment situations adopted by the Sixteenth ICLS in 1998. As a supplement to the Thirteenth ICLS Resolution, the Fourteenth ICLS (1987) endorsed Guidelines on the implications of employment promotion schemes on the measurement of employment and unemployment, and the Sixteenth ICLS (1998) Guidelines concerning the treatment in employment and unemployment statistics of persons on extended absences from work (ILO 2000).

[72] The working-age population is the population above the legal working age, but for statistical purposes it comprises all persons above a specified minimum age threshold for which an inquiry on economic activity is made. For reasons of international comparability, the working-age population is often defined as all persons aged 15 and older, but this may vary from country to country based on national laws and practices (some countries also apply an upper age limit).

unemployment using the information available on the unemployed. According to Bartholomew et al. (1995), the employment process can be described by a variety of measures of which totals, levels, rates and durations are the most common. Among them the unemployment rate is probably the best-known labour market measure.

Unemployment rate. The unemployment rate is defined as the ratio of total number of unemployed (n) to the total number of persons in the labour force (N) during the reference period. That is,

$$R = \frac{n}{N} = \frac{n}{e+n} \tag{5.2.1}$$

where e corresponds to the total number of persons in employment.

The overall unemployment rate for a country is a widely used measure of its untapped labour supply. When the unemployment rate is based on internationally recommended standards, it simply reflects the percentage of the labour force that does not have a job but is available and actively looking for work. Also, the unemployment rate can be considered as the most informative indicator of the labor market that reflects the overall performance of the labor market and the economy as a whole (ILO, 2019).

However, although the unemployment rate is a valuable indicator of the labor market, it is also inadequate in itself. It is very sensitive to seasonality and short-term fluctuations in the unemployment rate can be largely due to seasonal effects. Furthermore, the unemployment rate fails to provide any information on the quality of employment of those who do have a job, on the status of those excluded from the labour force and the conditions of the unemployed. It also does not provide a fully integrated measure of job underutilization, as it overlooks other forms of job underutilization, such as time-related underemployment and potential labour force (ILO, 2019). Therefore, in this context, other measures should complement the unemployment rate to fully assess labour underutilization, such as time-related underemployment and potential labor force indicators. Paul (1991) proposes a measure that takes into account both the intensity and distribution aspects of unemployment. This measure was based on the assumption that a person's misery in the labour force varies proportionately with the intensity of his unemployment, so a simple average of these intensities can be a good measure of unemployment.

In addition to the overall unemployment rate, the various labour surveys also provide unemployment rates for specific groups, defined by age, sex, occupation or industry. These unemployment rates are useful in identifying groups of workers and sectors most vulnerable to unemployment.

5.3 European Union Labour Force Survey

Labour force surveys are usually the preferred source of information for determining the unemployment rate. Such surveys can be designed to cover almost the entire population of a country and generally provide an opportunity to simultaneously measure the employed, unemployed and non-employed in a coherent framework. Also, population Censuses and other types of surveys for households can be used as data sources to calculate unemployment rates. However, the information obtained from these types of surveys may be less reliable, as they usually do not allow for detailed labor market research and job search activities of the respondents.

The European Union Labour Force Survey (EU LFS) is a rotating random sample survey covering the population in private households in currently 35 European countries. It is conducted by the National Statistical Institutes across Europe and is centrally processed by Eurostat (Eurostat, 2012). From 1998, the EU LFS has become a continuous quarterly survey. The main objective is to provide comparable information on employed, unemployed and inactive persons of working age (15 years and above) in European countries. The key issues of the LFS are (Eurostat, 2012):

- labour status
- employment characteristics of the main job
- hours worked
- second job
- previous work experience of person not in employment
- search for employment
- methods used during previous four weeks to find work
- main labour status
- education and training
- situation one year before survey
- income
- atypical work
- demographic background

The definitions of employment and unemployment used in the LFS closely follow the International Labour Organisation’s guidelines. The national statistical institutes are responsible for the selection of the sample, the preparation of the questionnaires, the conduct of the direct interviews between the households and the forwarding of the results to Eurostat in accordance with the requirements of the regulation. As national statistical institutes use the same concepts, definitions and classifications, all in accordance with the ILO guidelines, and record the same set of characteristics in each country, harmonized data are available at European level. In 2018, the quarterly LFS sample size across the EU was about 1.5 million individuals and covered all industries and occupations (Eurostat, 2020).

5.4 Small Area Estimation of Labour Force Indicators Using Fay-Herriot Model

Labour force surveys conducted in each country are the main source of national labour market information. However, they are not able to deliver direct estimates of unemployment with adequate precision for every local authority district because the sample size in many areas is insufficient (ONS, 2006). In the framework of the EURAREA project, small area estimators were developed and calculated in order to estimate the unemployment rates in small areas of various European countries that participated in the program. Real data from population Censuses or population registers were used to create a simulation database. The simulation settings make it possible to compare the area-level estimates, which were generated by the samples, with the true values from the population for the same areas. According to the results of the project, Eurostat encourages member states to adopt small area estimation methods for area sizes below (and possibly including) NUTS 3 (EURAREA, 2004).

Unemployment rates for small areas. Suppose a labour force population (employed and unemployed) of size N is divided into D small areas of size $N_1, N_2, \ldots, N_i, \ldots, N_D$ such that $\sum_{i=1}^{D} N_i = N$. Let,

$$y_{ij} = \begin{cases} 1, & \text{if unit } j \text{ in small area } i \text{ is unemployed} \\ 0, & \text{otherwise} \end{cases} \tag{5.4.1}$$

Then the unemployment rate R_i for the small area i $(i = 1,2, \ldots, D)$, is given by:

$$R_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij} \tag{5.4.2}$$

Direct estimator of unemployment rate for small areas. Suppose a random sample of size $n < N$ is drawn from the population according to a specified sampling design and $n_1, n_2, \dots, n_D$ is the sample size of the selected units in each small area $i = 1,2,\dots,D$. Let s_i be the set of units selected in the sample for the small area i and w_{ij} be the sampling weight of individual j from area i[73]. The basic Horvitz-Thompson direct estimator of the unemployment rate for the small area i ($i = 1,2,\dots,D$), is given by (D'Aló et al., 2012):

$$\hat{R}_i^{DIR} = \frac{1}{\hat{N}_i} \sum_{j \in S_i} w_{ij} y_{ij} \text{ , } i = 1,2,\dots,D \tag{5.4.3}$$

where

$$\hat{N}_i = \sum_{j \in S_i} w_{ij} \tag{5.4.4}$$

is the direct estimator of the total N_i of the *i*-th small area.

Indirect estimator of unemployment rate based on a Fay-Herriot model. According to formula (2.3.6), the Fay-Herriot area level model links the parameter of interest R_i for all the areas $i = 1,2,\dots,D$ through a linear model as:

$$R_i = \mathbf{x}_i{}^T \boldsymbol{\beta} + u_i \tag{5.4.5}$$

where $\boldsymbol{x_i} = (x_{1i}, x_{2i}, \dots, x_{pi})$ is a vector of covariates (area-specific auxiliary data) for domain i , $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ is the $p \times 1$ vector of regression coefficients and u_i's are area-specific random effects. The Fay-Herriot model assumes that $\hat{R}_i^{DIR}$ is design-unbiased, with:

$$\hat{R}_i^{DIR} = R_i + e_i, \; i = 1,2,\dots,D \tag{5.4.6}$$

where e_i is the sampling error associated with the direct estimate of each small area i. Combining (5.4.5) and (5.4.6) we obtain the linear mixed model:

$$\hat{R}_i^{DIR} = \boldsymbol{x}_i{}^T \boldsymbol{\beta} + u_i + e_i, \; i = 1,2,\dots,D \tag{5.4.7}$$

We assume that $u_i \overset{iid}{\sim} (0, \sigma_u^2)$ and $e_i \overset{ind}{\sim} (0, \psi_i)$, where the sampling variances ψ_i, $i = 1,2,\dots,D$, are supposed to be known.

The best linear unbiased predictor (BLUP) of $R_i = \mathbf{x}_i{}^T \boldsymbol{\beta} + u_i$ under model (5.4.7) can be expressed as a weighted combination of the direct $\hat{R}_i^{DIR}$ and the regression-synthetic estimators $\boldsymbol{x}_i{}^T \tilde{\boldsymbol{\beta}}$, that is (Guadarrama et al. 2014):

[73] It holds that $w_{ij} = \frac{1}{\pi_{ij}}$, where π_{ij} is the inclusion probability for individual j in area i.

$$\tilde{R}_i^{F-H} = \gamma_i \hat{R}_i^{DIR} + (1-\gamma_i)\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}, \ i = 1,2,\ldots,D \tag{5.4.8}$$

with weight

$$\gamma_i = \sigma_u^2 / (\sigma_u^2 + \psi_i) \tag{5.4.9}$$

and

$$\tilde{\boldsymbol{\beta}} = (\textstyle\sum_{i=1}^{D} \gamma_i \boldsymbol{x}_i \boldsymbol{x}_i^T)^{-1} \sum_{i=1}^{D} \gamma_i \boldsymbol{x}_i \hat{R}_i^{DIR} \tag{5.4.10}$$

is the weighted least squares estimator of $\boldsymbol{\beta}$.

In practice, the variance σ_u^2 of the area effects u_i is unknown and needs to be estimated. As already mentioned in paragraph 2.4.1, common estimation methods are maximum likelihood (ML) and restricted maximum likelihood (REML). Therefore, substituting $\hat{\sigma}_u^2$ for ${\sigma_u}^2$ in (5.4.8) we obtain the empirical best linear unbiased predictor (EBLUP) of R_i, denoted here as $\hat{R}_i^{F-H}$ and given by:

$$\hat{R}_i^{F-H} = \hat{\gamma}_i \hat{R}_i^{DIR} + \left(1 - \hat{\gamma}_i\right) \boldsymbol{x}_i{}^T \hat{\boldsymbol{\beta}}, \ i = 1,2,\ldots,D \tag{5.4.11}$$

where $\hat{\gamma}_i$ and $\hat{\boldsymbol{\beta}}$ are the values of γ_i and $\tilde{\boldsymbol{\beta}}$ when ${\sigma_u}^2$ is replaced by an estimator $\hat{\sigma}_u^2$.

Finally, for unsampled domains ($n_i = 0$) the unemployment rate is estimated using only auxiliary variables without sample data:

$$\hat{R}_i^{F-H} = \boldsymbol{x}_i{}^T \tilde{\boldsymbol{\beta}}, \ i = 1,2,\ldots,D \tag{5.4.12}$$

Estimates obtained in this way are called synthetic (Rao and Molina, 2015).

The Fay-Herriot model was used in the ESSnet project on SAE (2012d) to estimate unemployment rates France and Poland. In the case of France, the main objective was to experiment with different small area estimators (including the Fay and Herriot estimators) and compare them with the current estimator calculated by the French National Statistical Institute (Insee). Some of the auxiliary variables used were age, gender, level of education, people who spontaneously declare that they are looking for a job and people who declare that they leave their job on their own. In the case of Poland, the main goal was to estimate the percentage of unemployed people in the population of 15 and older at the NUTS3 level, that is the lower level of aggregation than presented in the Central Statistical Offices (CSO) publications. Seven estimators were applied, among them an EBLUP Fay and Herriot estimator.

Furthermore, the Office for National Statistics (ONS) of the United Kingdom has applied small area estimation methods in order to produce estimates of unemployment level and rate on the International Labour Organisation (ILO) definition for local authority districts and unitary authorities (LAD/UAs), (ONS, 2004). The diagnostic analysis confirmed that the models were well specified, stable and the assumptions were sound. Also, the models developed were robust, made the best use of the available data and the model-based estimates were plausible and informative to users.

López-Vizcaíno, Lombardía and Morales (2015) applied SAE methods to real data from the Spanish Labour Force Survey of Galicia in order to estimate labour force indicators like totals of employed and unemployed people and unemployment rates. The obtained small area estimates for all models developed were compared with the direct ones and they had lower mean squared errors, especially for counties with small sample size.

In addition, Omrani, Gerber and Bousch (2009) applied small area estimation techniques and proposed an Empirical Best Linear Unbiased Predictor (EBLUP) as an estimator in order to evaluate various indicators of unemployment. The major advantage of the approach proposed was the capability to combine, in an efficient way, several pieces of information from Censuses and registers, even in the case of incomplete data.

Finally, Meindl (2008) applied a simulation-based approach in order to evaluate small area methods for estimating unemployment rates. According to the simulation results model-based estimation methods provided acceptable results for estimating unemployment rates for the small domains under investigation in the simulation. Also, the estimation of unemployment rates at small area level was improved by using auxiliary information on unit-/area level.

6. Application to Unemployment: Estimating Unemployment in Greece Using Small Area Estimation Methods

6.1 Introduction

The main source of data for the collection and analysis of unemployment in the case of Greece is the Labour Force Survey[74] (LFS). NUTS 2 areas are the lowest geographic areas for which the LFS in Greece publishes estimates. LFS results are not published at a lower level (NUTS 3) because, due to small population and sample size, estimates in these areas have large sampling errors. This work aims to estimate the unemployment rate in Greece at a lower level of spatial aggregation than that used so far, that is at the level of sub regions-NUTS 3 (*Nomoi*), using the small area estimation methodology.

In line with the framework developed in the previous chapters, SAE methods were used to estimate the unemployment rate in Greece at NUTS 3 level at two different times, in 2009 (shortly before the start of the Greek financial crisis) and 2013 (during the crisis). Specifically, combining data from the EU-SILC survey[75] and the national Census of Greece, the Fay-Herriot model was applied to produce estimates for the unemployment rate of the Greek population. The underlying small area estimation model used unit level survey responses (EU-SILC survey) but area level auxiliary variables (Census data) due to the restrictions on linking unit level survey and Census data.

6.2 Research characteristics

The target parameters to be estimated were the unemployment rates for the years 2009 and 2013 in Greece. The Horvitz-Thompson direct estimator $\hat{R}_i^{DIR}$ (given by equation (5.4.3)) of the unemployment rate was derived for each small area i (*Nomoi*, $i =$

[74] The Labour Force Survey has produced estimates since 1981 (second quarter of the year). From 1998 onwards it has been a continuous quarterly survey. The main statistical objective of the Labour Force Survey is to divide the working age population (15 years and over) into three mutually exclusive and exhaustive groups persons in employment, unemployed and inactive persons. In addition, it collects information on demographic characteristics, on main job characteristics, on the existence and characteristics of a second job, on educational attainment, on participation in education, on previous working experience and on searching for a job (Hellenic Statistical Authority, 2013b).

$1,2, \ldots, 54$) for the years 2009 and 2013 using the unit level data from the EU-SILC survey 2009 and 2013 respectively.

The indirect estimates of the unemployment rate derived for the years 2009 and 2013 based on a Fay-Herriot model (as analyzed in paragraph 5.4). Area-specific auxiliary data from the Greek national Censuses of 2001 and 2011 were used to implement the Fay-Herriot model for the years 2009 and 2013 respectively. The variance σ_v^2 of the area-specific random effects was estimated using restricted maximum likelihood (REML) and then the empirical best linear predictor (EBLUP) $\hat{R}_i^{F-H}$ (given by equation (5.4.11)) of R_i (given by equation (5.4.2)) was obtained.

All computations were performed using the software R. Functions from the package *sae* (Molina and Marhuenda, 2015) as well as R functions which were developed during the ESSnet project in SAE (ESSnet, 2012b) were used. Specifically, from the package *sae* functions *eblupFH*() and *mseFH*() were used and the functions *mixed.area.sae*() and *diagnostic*() from the ESSnet project.

6.3 Data Sources

The model constructed in the present survey was based on data from the EU-SILC survey and the national Census of Greece. In particular, sample data for direct unemployment rate estimates at NUTS 3 level for the years 2009 and 2013 come from the EU –SILC 2009 and 2013[76] surveys respectively. The data of the auxiliary variables of the years 2009 and 2013 were derived from the national Greek Censuses of 2001 and 2011, respectively.

6.3.1 Sample data source. The initial goal of the research was to use microdata from the Labour Force Survey (LFS) in Greece for the years 2009 and 2013. However, for reasons of confidentiality and protection of personal data, the microdata was not provided. Thus, the micro data of EU-SILC that had already been provided were used. Nevertheless, it should be noted that there are some minor differences in the definition of unemployed persons between EU-SILC and the Census[77].

The variable PL031 (*Self-defined current economic status*) was used from the EU-SILC microdata. The unit in this variable is all current members of the household

[76] A detailed description of EU-SILC survey in Greece is given in paragraph 4.3.1
[77] Census definitions concerning the unemployed and the employed are given in paragraph 4.5

aged 16 and over and the values that the variable can take are the following (Eurostat, 2013):

- Employee working full-time
- Employee working part-time
- Self-employed working full-time (including family worker)
- Self-employed working part-time (including family worker)
- Unemployed
- Pupil, student, further training, unpaid work experience
- In retirement or in early retirement or has given up business
- Permanently disabled or/and unfit to work
- In compulsory military or community service
- Fulfilling domestic tasks and care responsibilities
- Other inactive person

According to the EU-SILC definitions (Eurostat, 2013) the concept of 'current' in this variable implies that any definitive changes in the activity situation are taken into account. It is also worth noting that this variable captures the individual's perception of his or her main activity at the moment. However, an individual's perception of his or her main activity may differ from the strict definitions used in the LFS[78] in Greece. For example, many people who consider themselves full-time students or housewives may be classified as employed in Greece LFS if they have a part-time job. Similarly, some people who consider themselves "unemployed" may not meet the LFS's strict criteria for taking active measures to find work and being available immediately.

Also, "work" means any work for pay or profit. Pay includes cash payments or "payment in kind" (payment in goods or services rather than money). Employees are defined as persons who work for a public or private employer and who receive

[78] According to LFS in Greece (Hellenic Statistical Authority, 2013b) the following are valid:

- Employed are persons aged 15 years or older, who during the reference week worked, even for just one hour, for pay or profit or they were working in the family business, or they were not at work but had a job or business from which they were temporarily absent.
- Unemployed are persons aged 15-74 who were without work during the reference week (they were not classified as employed), were currently available for work and were either actively seeking work in the past four weeks or had already found a job to start within the next three months.
- Inactive are those persons who are neither classified as employed nor as unemployed.
- Economically active population (labour force) are persons either employed or unemployed.
- Unemployment Rate is the ratio of unemployed divided by total labour force.

compensation in the form of wages, salaries, fees, gratuities, payment by results or payment in kind; non-conscripted members of the armed forces are also included (Eurostat, 2013). Based on the above, it should be emphasized that the results of the present survey on unemployment rates are difficult to compare with the corresponding results of the Labour Force survey. Actually, due to differences in data sets and definitions, the official employment rate from the LFS is consistently lower than the corresponding rate from the EU-SILC, but patterns are similar (Brandolini and Viviano, 2018).

In order to produce the direct estimates of the unemployment rate the following procedure was carried out. From the microdata of variable PL031, only those related to the labour force were used. That is, the unemployed and the employed. Then, a new survey binary variable was generated that took the value 1 when a person was classified as unemployed and the value 0 otherwise (employee).

6.3.2 Data of the auxiliary variables. National Greek Census data from the years 2001 and 2011 were used as area-specific auxiliary variables in order to apply the Fay-Herriot model and to produce indirect estimates of the unemployment rate in Greece for the years 2009 and 2013 respectively. The initial set of auxiliary variables used to estimate the unemployment rate in Greece is presented in Tables 4.5.1 and 4.5.2. The variable X12, which corresponds to the percentage of unemployed per prefecture, has been excluded from these, since the response variable is the same.

The initial set of auxiliary variables was selected based on the factors that seem to influence unemployment rate the most, according to the literature (Biagi and Bocconi, 2005; Grant, 2012; Hellenic Statistical Authority, 2009b and 2013b) as well as based on past researches to estimate unemployment using SAE methods (EURAREA, 2004; ESSnet, 2012d; D'Aló, Consiglio, Falorsi, Solari, Pratesi, Salvati, and Ranalli, 2012; López-Vizcaíno et al. 2015; Ugarte1, Goicoa, Militino and Sagaseta-Lopez, 2009; Meindl, 2008). Some of the key factors that are considered to affect unemployment are: Gender, age, educational attainment, disability or ill-health, foreign nationality, type of family. Specifically, in Greece, according to the data of LFS (Hellenic Statistical Authority, 2009b, 2013b) the groups with the highest unemployment rates were young people aged 15-29 years old, females, persons who have not attended school and persons with foreign nationality.

Based on the above theoretical framework and the availability of Census data, the main types of variable selected include variables related to demographic characteristics, socioeconomic status and educational attainment.

6.4 Horvitz-Thompson (H-T) direct estimates of the unemployment rate in Greece for the years 2009 and 2013

In order to implement the Fay and Herriot model, two components are required: a sampling model for the direct estimates and a linking model for the parameters of interest. The direct estimator of a small area uses only the sample data from the target small area and in the present study derived from the EU-SILC survey in Greece for the years 2009 and 2013. The formulas used to calculate the direct estimates of unemployment rate for the years 2009 and 2013 are given in the Tables 6.5.1 and 6.5.2 respectively. The results of the direct estimates as well as the corresponding variances and coefficients of variations (CV) for the years 2009 and 2013 are given in Tables A15 and A16 in the Appendix. It should be noted that some prefectures dropped out of the estimation process as the numerical value of the direct estimate of unemployment rate for the specific prefectures was zero. For the year 2009 these were the prefectures of Lefkada (300024), Arta (300031) and Samos (300084) and for the year 2013 the prefecture of Lefkada.

Table 6.5.1 Summary formulas for the Horvitz-Thompson (H-T) direct estimators of unemployment rate for the year 2013

N_i : labour force population size of the *i*-th prefecture n_i: sample size of the *i*-th prefecture s_i: set of units selected in the sample for the *i*-th prefecture w_{ij}: sampling weight for the *j* unit in the *i*-th prefecture $i = 1, \ldots, 53$
Parameters to be estimated
• The unemployment rate $R_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}, \quad i = 1, \ldots, 53$ where $y_{ij} = \begin{cases} 1, & \text{if unit } j \text{ in small area } i \text{ is unemployed} \\ 0, & \text{otherwise} \end{cases}$
H-T direct estimators
• $\hat{R}_i^{DIR} = \frac{1}{\hat{N}_i} \sum_{j \in S_i} w_{ij} y_{ij}, \; i = 1, \ldots, 53$ where $\hat{N}_i = \sum_{j \in s_i} w_{ij}$ is the direct estimator of the population size N_i of the *i*-th prefecture.

Table 6.5.2 Summary formulas for the Horvitz-Thompson (H-T) direct estimators of unemployment rate for the year 2009

N_i : labour force population size of the *i*-th prefecture n_i: sample size of the *i*-th prefecture s_i: set of units selected in the sample for the *i*-th prefecture w_{ij}: sampling weight for the *j* unit in the *i*-th prefecture $i = 1, \ldots, 51$
Parameters to be estimated
• The unemployment rate $R_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}, \; i = 1, \ldots, 51$ where $y_{ij} = \begin{cases} 1, & \text{if unit } j \text{ in small area } i \text{ is unemployed} \\ 0, & \text{otherwise} \end{cases}$
H-T direct estimators
• $\hat{R}_i^{DIR} = \frac{1}{\hat{N}_i} \sum_{j \in S_i} w_{ij} y_{ij}, \; i = 1, \ldots, 51$ where $\hat{N}_i = \sum_{j \in s_i} w_{ij}$ is the direct estimator of the population size N_i of the *i*-th prefecture.

6.5 Model selection

In order to build the optimal small area model for estimating unemployment rate in Greece for the years 2009 and 2013, a three-phase variable selection process was performed. This process is analyzed and described in detail in paragraph 4.6. In brief:

- In the first phase a correlation matrix[79] among the auxiliary variables (the initial set of auxiliary variables as presented in Tables 4.5.1 and 4.5.2) was analyzed. From this analysis a subset of variables was excluded based on whether they were correlated with other covariates, how many other covariates, and how high the correlation was.
- In the second phase a forward stepwise procedure was used starting with a null model, i.e., just an intercept and then covariates were added one-by-one until there was no improvement in terms of a selection criterion. That is, of the models produced, the acceptable ones were those where: all included covariates were significant (p-values of the significance of each coefficient less than 10%), and no others were significant enough to enter the model and the sign next to each variable (that is, the sign of estimated model coefficients - beta parameter) was justified according to the knowledge of the analyzed phenomena from the literature. For example, it is expected that when the percentage of young people increases, so do the levels of unemployment rate, so the symbol next to the variable corresponding to young people should be positive.
- In the third phase the final model for each year was chosen according to three information criteria. These criteria were the Akaike information criterion (AIC), Bayesian information criterion (BIC) and conditional Akaike Information Critieria (cAIC).

The models that were created from the first two phases are presented in Tables 6.5.3 and 6.5.4 below. In these tables the estimated model coefficients, the p-value of the significance of each coefficient and the information criteria for each model are presented.

[79] The correlation matrices for both the variables in Table 4.4.1 and the variables in Table 4.4.2 are presented in the Appendix in Tables A1 and A2.

Table 6.5.3 Coefficients (estcoef), estimated variance of the area random effects (refvar) and information criteria (goodness) for the selected models of 2013 for estimating unemployment rate using the EBLUP F-H model

model 1

X7: People aged 29-49 years old.
X10: Women
X28: Five or more members
X30: Without employed member

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-2.396326	0.6909764	-3.468029	0.0005242908
X7	3.230126	1.0782123	2.995817	0.0027371100
X10	2.301769	1.1778252	1.954253	0.0506712755
X28	1.189432	0.5067696	2.347086	0.0189208948
X30	1.435298	0.4662890	3.078130	0.0020830402

refvar
[1] 0.001238698

goodness

loglike	AIC	BIC	KIC	cAIC
67.2673	-122.5346	-110.7129	-116.5346	-130.041

model 2

X7: People aged 29-49 years old.
X13: Inactive people
X31: Single-parent families
X32: Without car

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-3.169502	0.6243249	-5.076687	3.840726e-07
X7	4.556718	1.0164322	4.483051	7.358324e-06
X13	3.018514	0.5763034	5.237717	1.625754e-07
X31	1.687084	0.5091799	3.313336	9.219030e-04
X32	0.353888	0.1871061	1.891376	5.857413e-02

refvar
[1] 0.000115993

goodness

loglike	AIC	BIC	KIC	cAIC
74.5912	-137.1824	-125.3607	-131.1824	-139.7346

model 3

X8: People aged over 50 years old
X25: Rented dwellings
X31: Single-parent families
X32: Without car

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	0.428521	0.1299905	3.296555	9.787830e-04
X8	-1.327065	0.3263219	-4.066737	4.767590e-05
X25	-1.062947	0.2301708	-4.618077	3.873132e-06
X31	2.327597	0.4877466	4.772144	1.822754e-06
X32	0.789277	0.2006759	3.933093	8.385985e-05

refvar
[1] 0.0003027123

goodness

loglike	AIC	BIC	KIC	cAIC
73.16186	-134.32373	-122.50197	-128.32373	-137.6856

model 4

X10: Women
X25: Rented dwellings
X27: Three or more children
X31: Single-parent families

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-1.2912668	0.5696373	-2.266823	0.023401050
X10	2.6838457	1.1533692	2.326961	0.019967318
X31	1.6277893	0.5678188	2.866741	0.004147227
X25	-0.4991047	0.1940051	-2.572636	0.010092717
X27	1.0708449	0.5435931	1.969938	0.048845422

refvar
[1] 0.0008604319

goodness

loglike	AIC	BIC	KIC	cAIC
67.85806	-123.71612	-111.89436	-117.71612	-129.5386

model 5

X2: Medium education
X10: Women
X25: Rented dwellings
X27: Three or more children

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-1.8018420	0.5776757	-3.119123	0.001813899
X10	3.6202660	1.1229927	3.223766	0.001265166
X2	0.7994718	0.2883976	2.772116	0.005569314
X25	-0.7535128	0.2402422	-3.136471	0.001709943
X27	1.1586314	0.5497819	2.107438	0.035079597

refvar
[1] 0.0009430976

goodness

loglike	AIC	BIC	KIC	cAIC
67.63646	-123.27293	-111.45118	-117.27293	-129.4783

model 6

X9: People aged over 65 years old
X10: Women
X25: Rented dwellings
X31: Single-parent families

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.7315052	0.5569981	-1.313299	0.189082097
X10	2.0935991	1.1604514	1.804125	0.071211742
X31	1.6750310	0.6036121	2.775012	0.005519965
X25	-0.8231626	0.2589466	-3.178889	0.001478407
X9	-0.6372238	0.3565757	-1.787065	0.073927054

refvar
[1] 0.001272807

goodness

loglike	AIC	BIC	KIC	cAIC
67.39972	-122.79944	-110.97769	-116.79944	-130.528

model 7

X7: People aged 29-49 years old.
X8: People aged over 50 years old
X13: Inactive people
X30: Without employed member
X31: Single-parent families

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-1.911907	0.7277065	-2.627305	8.606404e-03
X7	2.628115	1.1900015	2.208497	2.720962e-02
X8	-1.096459	0.4475997	-2.449643	1.429979e-02
X13	1.836318	0.7656825	2.398276	1.647246e-02
X30	1.245891	0.5791960	2.151070	3.147067e-02
X31	2.481232	0.5232048	4.742373	2.112297e-06

refvar
[1] 0.0001524493

goodness

loglike	AIC	BIC	KIC	cAIC
76.18454	-138.36908	-124.57703	-131.36908	-141.119

model 8

X7: People aged 29-49 years old.
X9: People aged over 65 years old
X13: Inactive people
X30: Without employed member
X31: Single-parent families

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-2.146445	0.7228700	-2.969337	2.984431e-03
X7	2.668330	1.2669441	2.106115	3.519436e-02
X9	-1.246760	0.5985704	-2.082897	3.726063e-02
X13	2.055474	0.7692800	2.671945	7.541305e-03
X30	1.101700	0.5811958	1.895575	5.801632e-02
X31	2.263884	0.5167824	4.380730	1.182825e-05

refvar
[1] 0.000244439

goodness

loglike	AIC	BIC	KIC	cAIC
75.29561	-136.59122	-122.79918	-129.59122	-139.7598

Note: Models created using auxiliary variables of the 2011 national Greek Census and EUSILC data 2013

Table 6.5.4 Coefficients (estcoef), estimated variance of the area random effects (refvar) and information criteria (goodness) for the selected models of 2009 for estimating unemployment rate using the EBLUP F-H model

model 1

X8: People aged over 50 years old
X13: Inactive people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.09954857	0.1362138	-0.7308257	0.46488563
X8	-0.60385355	0.2418926	-2.4963707	0.01254714
X13	0.72922496	0.3168041	2.3018164	0.02134553

refvar
[1] 0.001160165

goodness

loglike	AIC	BIC	KIC	cAIC
77.26483	-146.52966	-138.80235	-142.52966	-167.7872

model 2

X9: People aged over 65 years old
X13: Inactive people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.2408761	0.1435903	-1.677524	0.093440099
X9	-0.8687084	0.2672391	-3.250679	0.001151298
X13	0.8853959	0.3148524	2.812099	0.004921941

refvar
[1] 0.001011908

goodness

loglike	AIC	BIC	KIC	cAIC
79.13552	-150.27104	-142.54374	-146.27104	-169.7223

model 3

X2: Medium education
X6: People aged under 15 years old
X13: Inactive people

estcoef

	beta	std.error	tvalue	pvalue
(Intercept)	-0.4893503	0.2262080	-2.163276	0.03051995
X6	1.2108118	0.6028780	2.008386	0.04460229
X2	0.2365761	0.1424820	1.660393	0.09683545
X13	0.5446683	0.2799618	1.945509	0.05171373

refvar
[1] 0.001200034

goodness

loglike	AIC	BIC	KIC	cAIC
77.15872	-144.31743	-134.65830	-139.31743	-166.8465

Note: Models created using auxiliary variables of the 2001 national Greek Census and EUSILC data 2009

The results of the information criteria from the third phase for each model in Tables 6.5.3 and 6.5.4 are presented in Tables 6.5.5 and 6.5.6.

Table 6.5.5 Information criteria AIC, BIC, cAIC for the selected models for estimating unemployment rate for the year 2013

Final Models	**Information Criteria**		
	AIC	BIC	cAIC
model 1	-122.5346	-110.7129	-130.041
model 2	-137.1824	**-125.3607**	-139.7346
model 3	-134.32373	-122.50197	-137.6856
model 4	-123.71612	-111.89436	-129.5386
model 5	-123.27293	-111.45118	-129.4783
model 6	-122.79944	-110.97769	-130.528
model 7	**-138.36908**	-124.57703	**-141.119**
model 8	-136.59122	-122.79918	-139.7598

Note: The figures displayed in black denote the best values for AIC, BIC and cAIC.
Models created using auxiliary variables of the 2011 national Greek Census and EUSILC data 2013.

Table 6.5.6 Information criteria AIC, BIC, cAIC for the selected models for estimating unemployment rate for the year 2009

Final Models	**Information Criteria**		
	AIC	BIC	cAIC
model 1	-146.52966	-138.80235	-167.7872
model 2	**-150.27104**	**-142.54374**	**-169.7223**
model 3	-144.31743	-134.65830	-166.8465

Note: The figures displayed in black denote the best values for AIC, BIC and cAIC.
Models created using auxiliary variables of the 2001 national Greek Census and EUSILC data 2009.

According to the above tables the smallest value for all three information criteria corresponds to model 2 (Table 6.5.6) for the year 2009. For the year 2013 the smallest value for AIC and cAIC corresponds to model 7 (Table 6.5.5) while the smallest value for BIC corresponds to model 2. As already mentioned in paragraph 2.6.1 the BIC penalizes models with a greater number of variance parameters more than the AIC does. As a result, the two criteria may lead to different models as here. Therefore, cAIC which is well-suited for small area estimation[80], was used for the selection of the final model. Finally, model 2 from the year 2009 and model 7 from the year 2013 were selected as the final models for carrying out the estimation process.

6.6 Model Diagnostics and Evaluation

Once one model has been selected, it is necessary to assess the quality of the fit of the model and the performance of estimation method. The procedure followed in order to evaluate the final model, described in detail in paragraph 4.7, includes:

- Bias diagnostic
- Goodness of fit diagnostic
- Coverage diagnostic
- Residual analysis
- Gain in precision indexes

The numerical values of the EBLUP Fay and Herriot estimates as well as the corresponding mean square errors (MSE), the coefficients of variation (CV), the GIP1 and GIP2 ratios are given in Tables A17, A18, A19 and A20 in the Appendix.

[80] Details are given in paragraph 2.6.1

6.6.1 Model Diagnostics for the estimation of unemployment rate in Greece for the year 2013. For the estimation of unemployment rate model 7 (Table 6.5.3) was selected. In model 7 there were five covariates: Percentage of people per prefecture aged 29-49 years old (X7), percentage of people per prefecture aged over 50 years old (X8), percentage of inactive people per prefecture (X13), percentage of nuclear families per prefecture without an employed member (X30), percentage of single-parent families per prefecture (X31). Beta (β) parameters that corresponds to variables X7, X13, X30 and X31 have a positive sign (Table 6.6.1), which means that the increase of one of those four indicators in an area is associated with an increase of the unemployment rate in this unit. Beta parameter that corresponds to variable X8 have a negative sign (Table 6.6.1), which means that the increase of people aged over 50 years old in a prefecture affects a decrease of the unemployment rate in this unit.

Table 6.6.1 Coefficients for the final selected model for estimating unemployment rate for the year 2013 using F-H model

Unemployment rate (R_i)				
	beta	std.error	tvalue	pvalue
(Intercept)	-1.911907	0.7277065	-2.627305	8.606404e-03
X7 people aged 29-49 years old	2.62815	1.1900015	2.208497	2.720962e-02
X8 people aged over 50 years old	-1.096459	0.4475997	-2.449643	1.429979e-02
X13 inactive people	1.836318	0.7656825	2.398276	1.647246e-02
X30 without employed member	1.245891	0.5791960	2.151070	3.147067e-02
X31 single-parent families	2.481232	0.5232048	4.742373	2.112297e-06

i) In order to examine whether a substantial bias exists, the graphical diagnostic suggested by Brown et al. (2001) (analyzed in section 4.7 (i)) was produced and illustrated in Figure 6.6.1 for the estimate of unemployment rate. In this figure the bisector is shown in black, whereas the regression line is red. Also, a goodness of fit diagnostic proposed by Brown et al. (2001) (analyzed in section 4.7 (ii)) was produced[81], and the results are given in Table 6.6.2.

As noted in Figure 6.6.1 the intercept of the linear regression is $\beta_0 = 0.02203$ and the parameter for the slope is $\beta_1 = 0.95996$. The intercept is close to zero and the slope estimate does not differ much from 1. It seems that the cloud of points spreads along the line Y=X, so there is a strong assumption of lack of bias. The goodness of fit diagnostic, as shown in Table 6.6.2, accept null hypothesis that Fay-Herriot estimates are close to the direct estimates when the direct estimates are good. Therefore, it appears that the model for estimating unemployment rate is unbiased.

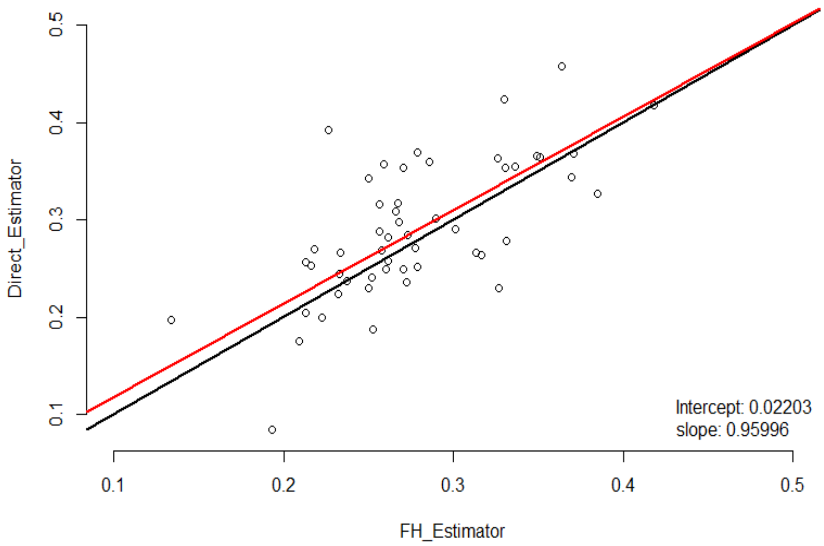

Figure 6.6.1 Relation between direct estimates and Fay-Herriot estimates of the unemployment rate in Greece for the year 2013

[81] For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used. This R function was developed for the application of model bias diagnostic proposed by Brown et al. (2001).

Table 6.6.2 The values of the empirical Wald test (W), of the theoretical χ^2 (c_alfa1), the p-value and the test result for the Fay and Herriot model estimator of the unemployment rate in Greece for the year 2013

method	W	c_alfa1	p-value	results
eblup.area	29.76615	69.83216	0.005624898	Accept H0: E(Direct estimates) = Model based Estimates

ii) In order to evaluate the validity of the confidence intervals generated by the Fay and Herriot model a coverage diagnostic (analyzed in section 4.7 (iii))[82] was used. The results are given in Table 6.6.3. The null hypothesis that the overlap is 95% is accepted. This means that the confidence intervals generated by the Fay and Herriot model are valid. Also, the numerical values of the confidence intervals for the Fay and Herriot estimator are given in Table A17 in the Appendix. An illustration of the results is presented in Figure 6.6.2.

Table 6.6.3 The values of the empirical z, of the theoretical z (z_teo) , the p-value, the overlapped areas, the overlap rate (f_sovrap) and the result of the test for the Fay and Herriot model estimator of the unemployment rate in Greece for the year 2013

method	z	z_teo	p_value	overlap	f_sovrap	results
eblup.area	1.670172	1.96	0.09488539	53	1.000000	Accept H0: The overlap is 95%

[82] For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used.

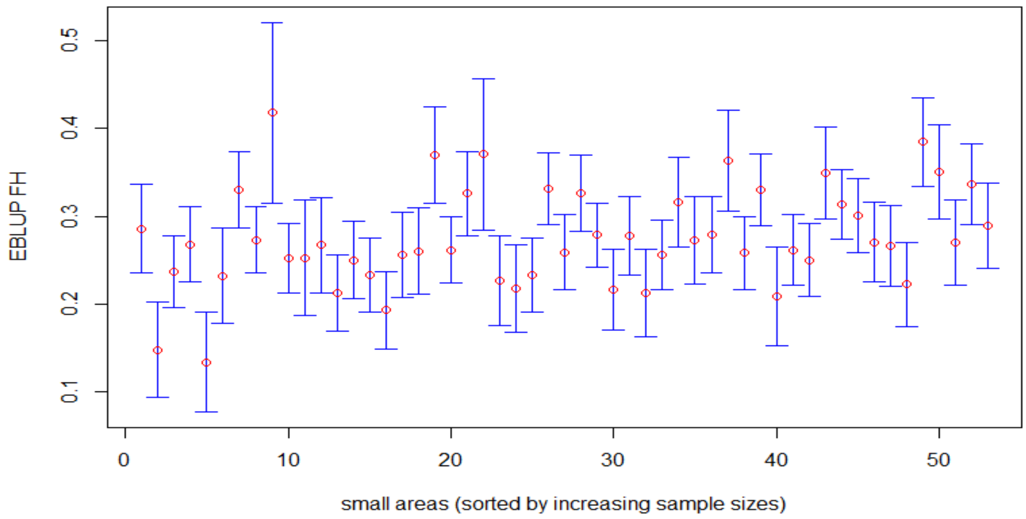

Figure 6.6.2 Confidence intervals for the Fay and Herriot estimators of the unemployment rate in Greece for the year 2013 (sorted by increasing sample sizes)

iii) To test the hypothesis of normal distribution of the sampling errors a Q-Q plot for the standardized residuals (Figure 6.6.3), a Shapiro-Wilk test for normality (Table 6.6.4), a plot of Fay-Herriot model versus standardized residuals as well as a histogram of the residuals were produced (Figure 6.6.4) (analyzed in section 4.7 (iv))[83].

Based on Figure 6.6.3 the normal Q-Q plot of standardized residuals shows that standardized residuals are normally distributed since they lie on a straight line, even if there are some outliers. The Shapiro-Wilk test for normality confirms the above finding since with the p-value = 0.6029 we cannot reject the null hypothesis of normality of the sampling errors. The hypothesis of normality is also confirmed from the histogram of the residuals. Furthermore, in the plot of Fay-Herriot model estimates versus standardized residuals there is not an obvious pattern in those residuals. Therefore, it seems that the assumption of constant variance of the sampling errors is satisfied.

[83] All computations were performed using software R.

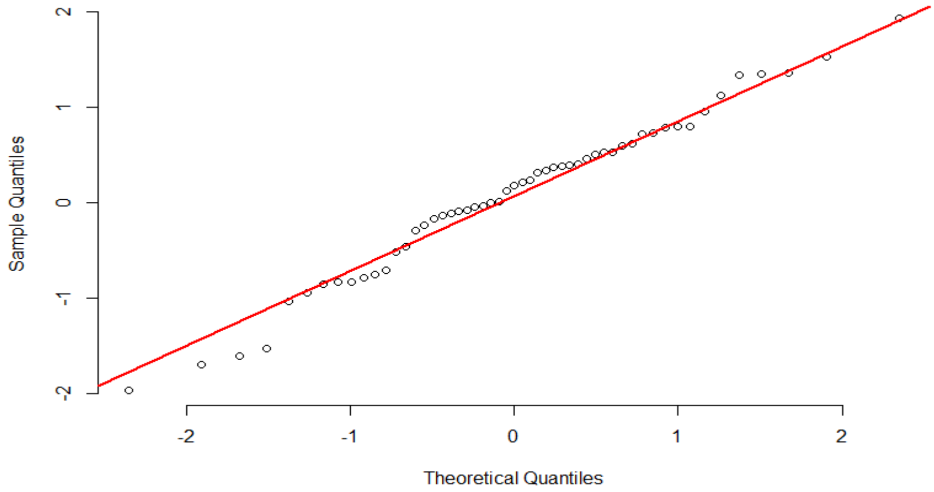

Figure 6.6.3 Normal Q-Q plot of standardized residuals of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2013

Table 6.6.4 Shapiro-Wilk test for normality of the sampling errors of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2013

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.98202	p-value=0.6029

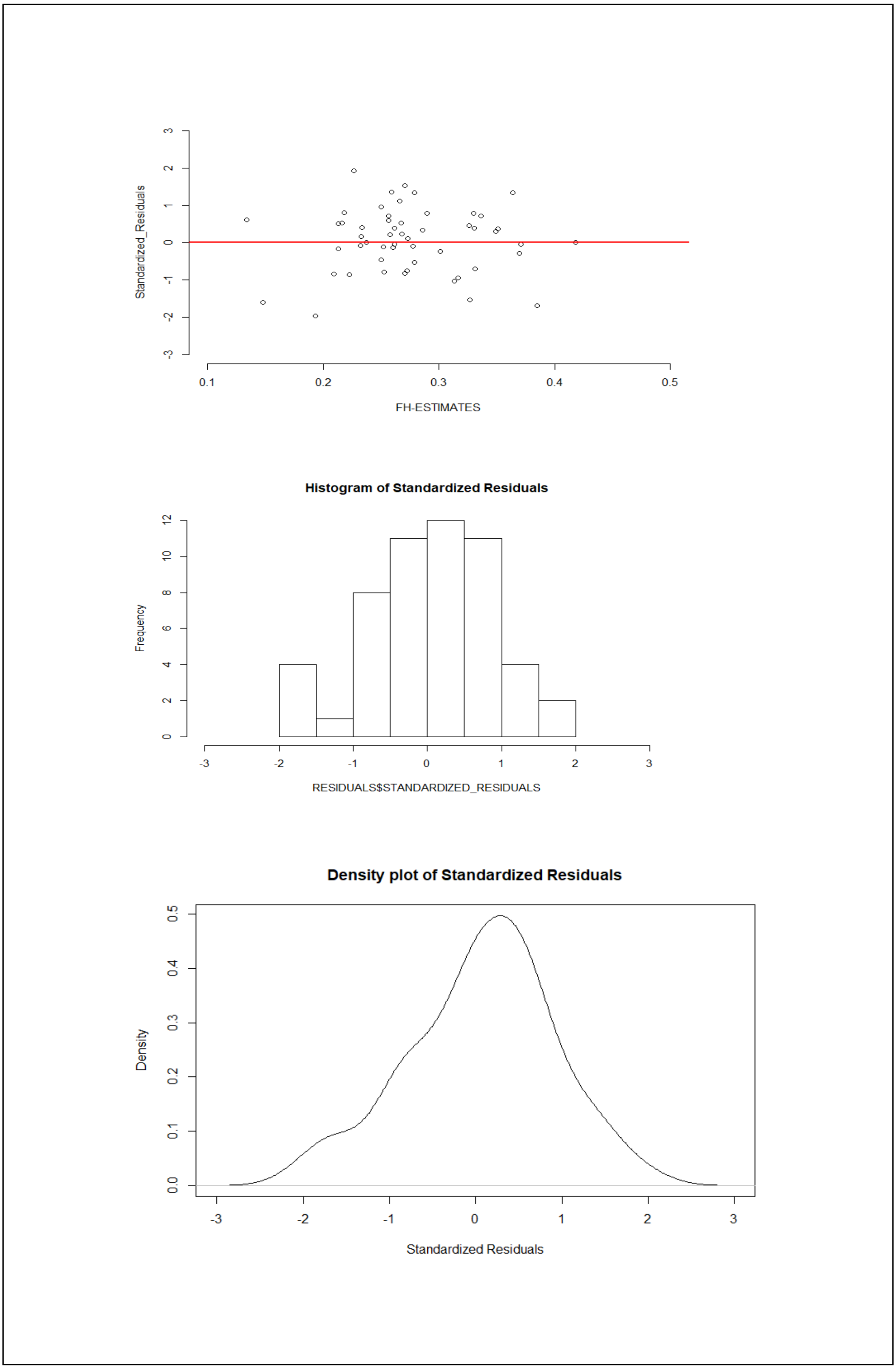

Figure 6.6.4 Residual distribution of the Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2013

iv. A similar procedure as for the sampling errors was followed to check the hypothesis of normality of the random effects. Based on Figure 6.6.5 there are some outliers in the Q-Q plot but the Shapiro-Wilk test (Table 6.6.5) as well as the histogram of random effects (Figure 6.6.6) confirm the hypothesis of normality. Also, the cloud of points in the plot of Fay-Herriot model versus random effects has no obvious pattern. In conclusion the hypothesis of normality of random effects seems to be satisfied.

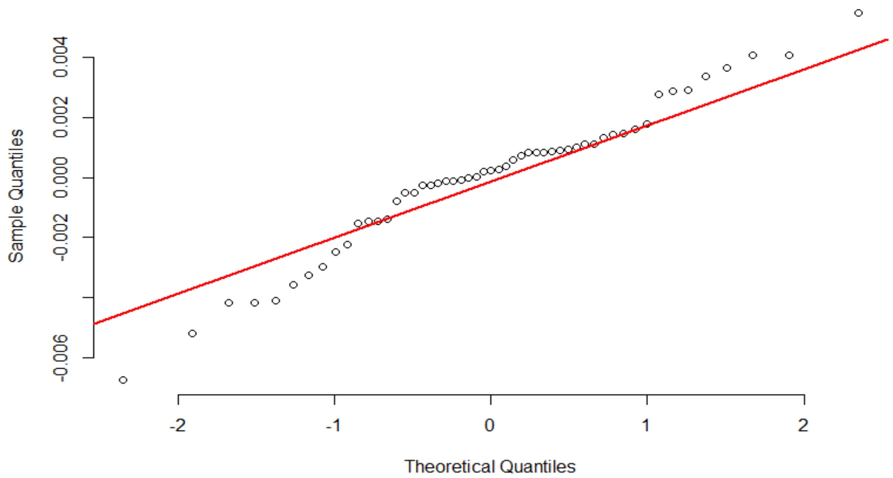

Figure 6.6.5 Normal Q-Q plot of random effects of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2013

Table 6.6.5 Shapiro-Wilk test for normality of the random effects of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2013

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.9681	p-value=0.1669

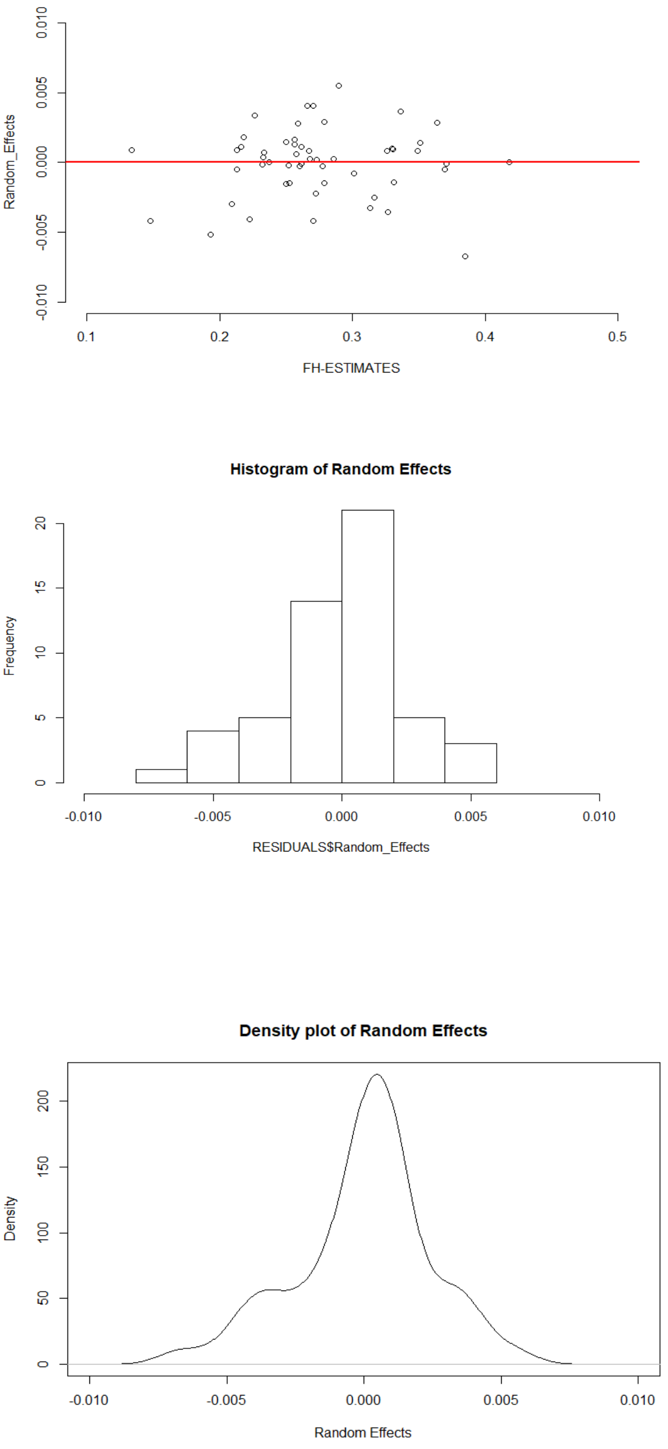

Figure 6.6.6 Random effects distribution of the Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2013

v) In order to evaluate the models, their quality should be checked. As already described in sections 2.6.2 and 7.4 (v and vi) coefficient of variation (CV) and mean square error (MSE) have been used as quality measures. Figures 6.6.7 and 6.6.8 illustrates the standard errors and the coefficients of variation (CV) of direct and Fay-Herriot estimates as well as the precision gain index GIP1 and GIP2. Detailed information about direct estimates, Fay and Herriot estimates and the corresponding standard errors, coefficients of variation (CV), precision gain index GIP1 and GIP2 can be found in the Appendix in Table A19.

Based on Figure 6.6.7 there is one area (prefecture of Athens, code: 300101) in which the CV and standard error of the direct estimate is less than the CV and MSE of the Fay and Herriot estimate. This is something to be expected since this area has a large enough sample size (n=1672) to give an accurate direct estimate. Nevertheless, there is overall a clear gain of precision when using the Fay-Herriot estimators instead of the direct estimators. This gain is seen in both the standard error ratio (GIP2) and the estimated MSE ratio (GIP1). The improvement in precision gain tends to be greater for areas with a smaller sample size. Indeed, areas with a small sample size such as the prefectures of Grevena (code:300051, n=9, GIP1=8.8, GIP2=7), Samos (code:300084, n=11, GIP1=5.17, GIP2=5.17) and Chios (code:300085, n=20, GIP1=5.89, GIP2=5.31) have a large gain in precision. For example, in Grevena the standard error of the Fay-Herriot estimate was reduced 8.8 times and the CV 7 times in relation to the direct estimate. This is evident as the direct estimator is likely to be more unstable in areas with small sample size. Summarizing, the application of small area estimation approaches achieved an overall significant efficiency gain for the estimation of unemployment rate in Greece for the year 2013.

National statistical offices usually establish a maximum publishable CV. As pointed out by Molina and Marhuenda (2015) and ONS (2004) estimates are considered precise and are suitable for publication when the majority of CV are below 20%. For these data the estimated CVs of direct estimators of unemployment rate exceeded the level of 20% for 34 (out of the 53) domains while those of the EBLUP F-H estimators exceeded this level for only one domain.

In conclusion, on the one hand the assumptions of the Fay-Herriot model seem to be satisfied for the unemployment rate in Greece for the year 2013 and on the other there is a clear overall precision gain of the application of the Fay-Herriot model. Finally, the majority of CV of headcount ratio are below 20% (only one exception).

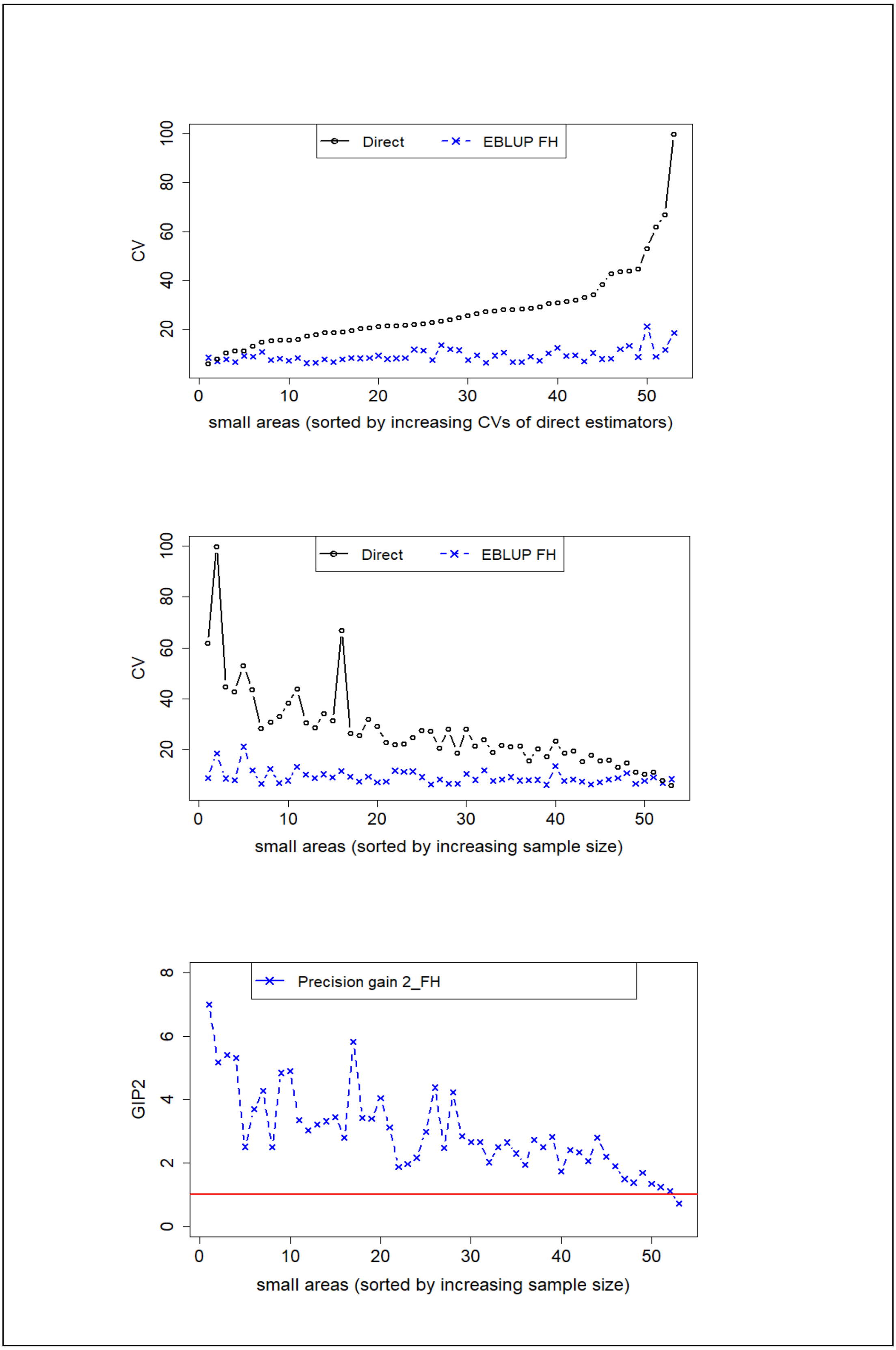

Figure 6.6.7 Coefficients of variation for the Direct and Fay-Herriot estimator of the unemployment rate in Greece for the year 2013 (sorted by increasing sample size and sorted by increasing CVs of direct estimator) and gain in precision index (GIP2)

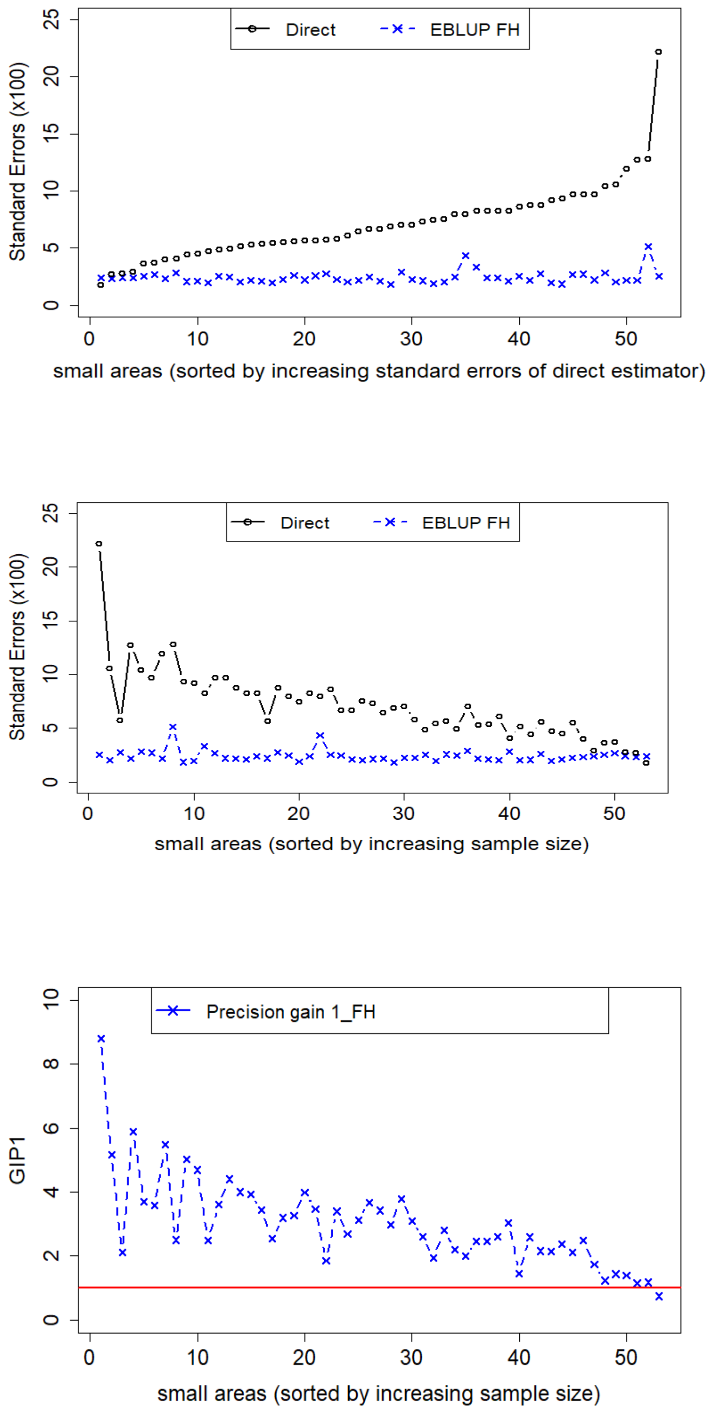

Figure 6.6.8 Standard errors for the Direct and Fay-Herriot estimator of the unemployment rate in Greece for the year 2013 and gain in precision index (GIP1) sorted by increasing sample size

6.6.2 Model Diagnostics for the estimation of unemployment rate in Greece for the year 2009. For the estimation of unemployment rate model 2 (Table 6.5.4) was selected. In model 2 there were two covariates: Percentage of people per prefecture over 65 years old (X9) and percentage of inactive people per prefecture (X13). Beta (β) parameter that corresponds to variable X13 have a positive sign (Table 6.6.6), which means that the increase of inactive people is associated with an increase of the unemployment rate in this unit. Beta parameter that corresponds to variable X9 have a negative sign (Table 6.6.6), which means that the increase of people aged over 65 years old in a prefecture affects a decrease of the unemployment rate in this unit.

Table 6.6.6 Coefficients for the final selected model for estimating the unemployment rate for the year 2009 using F-H model

Unemployment rate (R_i)				
	beta	std.error	tvalue	pvalue
(Intercept)	-0.2408761	0.1435903	-1.677524	0.093440099
X9 people aged over 65 years old	-0.8687084	0.2672391	-3.250679	0.001151298
X13 inactive people	0.8853959	0.3148524	2.812099	0.004921941

i) In order to examine whether a substantial bias exists, the graphical diagnostic suggested by Brown et al. (2001) (analyzed in section 4.7 (i)) was produced and illustrated in Figure 6.6.9 the estimate of unemployment rate. In this figure the bisector is shown in black, whereas the regression line is red. Also, a goodness of fit diagnostic proposed by Brown et al. (2001) (analyzed in section 4.7 (ii)) was produced[84], and the results are given in Table 6.6.7.

As noted in Figure 6.6.9 the intercept of the linear regression is $\beta_0 = 0.03765$ and the parameter for the slope is $\beta_1 = 0.87064$. The intercept is close to zero, but slope estimates are slightly different from 1. There seems to be a slight disparity from the line Y=X, so there is a possible bias. However, the goodness of fit diagnostic, as shown in Table 6.6.7, accept null hypothesis that the Fay-Herriot estimates are close to

[84] For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used. This R function was developed for the application of model bias diagnostic proposed by Brown et al. (2001).

the direct estimates when the direct estimates are good. Therefore, it appears that the model for estimating unemployment rate is unbiased.

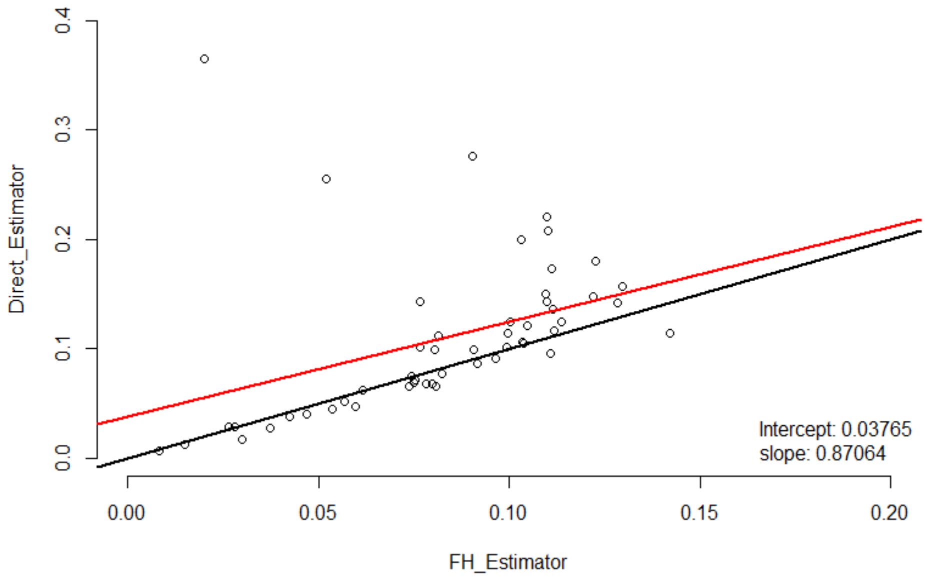

Figure 6.6.9 Relation between direct estimates and Fay-Herriot estimates of the unemployment rate in Greece for the year 2009

Table 6.6.7 The values of the empirical Wald test (W), of the theoretical χ^2 (c_alfa1), the p-value and the test result for the Fay and Herriot model estimator of the unemployment rate in Greece for the year 2009

method	W	c_alfa1	p-value	results
eblup.area	18.50886	67.50481	1.366321e-05	Accept H0: E(Direct estimates) = Model based Estimates

ii) In order to evaluate the validity of the confidence intervals generated by the Fay and Herriot model a coverage diagnostic (analyzed in section 4.7 (iii))[85] was used. The results are given in Table 6.6.8. The null hypothesis that the overlap is 95% is accepted. This means that the confidence intervals generated by Fay and Herriot model are valid. Also, the numerical values of the confidence intervals for the Fay and Herriot estimator

[85] For this goodness of fit statistic, the R function *diagnostic* developed under the ESSnet project in SAE (ESSnet, 2012b) was used.

are given in Table A18 in the Appendix. An illustration of the results is presented in the Figure 6.6.10.

Table 6.6.8 The values of the empirical z, of the theoretical z (z_teo), the p-value, the overlapped areas, the overlap rate (f_sovrap) and the result of the test for the Fay and Herriot model estimator of the unemployment rate in Greece for the year 2009

method	z	z_teo	p_value	overlap	f_sovrap	results
eblup.area	1.638356	1.96	0.1013474	51	1.000000	Accept H0: The overlap is 95%

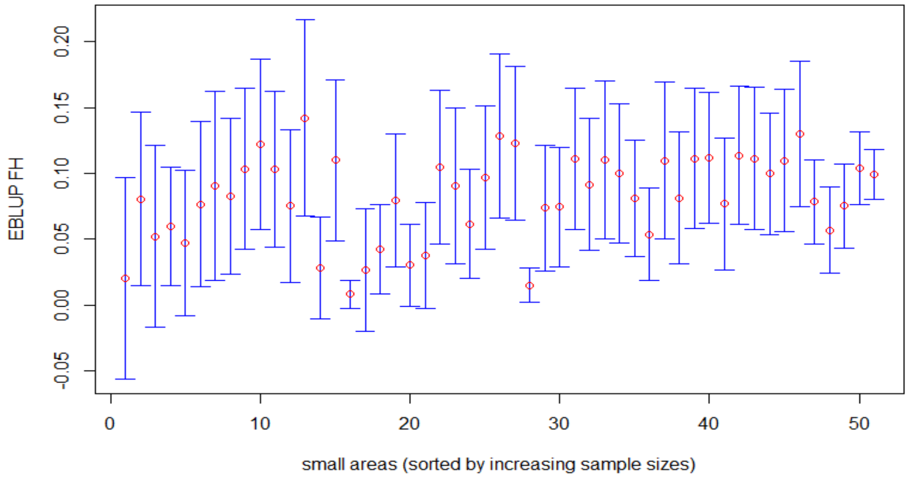

Figure 6.6.10 Confidence intervals for the Fay and Herriot estimators of the unemployment rate in Greece for the year 2009 (sorted by increasing sample sizes)

iii) To test the hypothesis of normal distribution of the sampling errors a Q-Q plot for the standardized residuals (Figure 6.6.11), a Shapiro-Wilk test for normality (Table 6.6.9), a plot of Fay-Herriot model versus standardized residuals as well as a histogram of the residuals were produced (Figure 6.6.12) (analyzed in section 4.7 (iv))[86].

Based on Figure 6.6.12, there appears to be a slight slope in both the histogram and the density plot of standardized residuals. Nevertheless, the normal Q-Q plot (Figure 6.6.11) shows that the standardized residuals are normally distributed since they lie on a straight line. The Shapiro-Wilk test for normality confirms the above finding since with the p-value= 0.8433 we cannot reject the null hypothesis of normality of the sampling errors. Furthermore, in the plot of Fay-Herriot model estimates versus

[86] All computations were performed using software R.

standardized residuals it seems that there is a notable but not so clear pattern (a connection between the residuals and Fay-Herriot estimates). This could have an impact on the calculation of confidence intervals as the assumption of the constant variance of the sampling errors is not clearly satisfied.

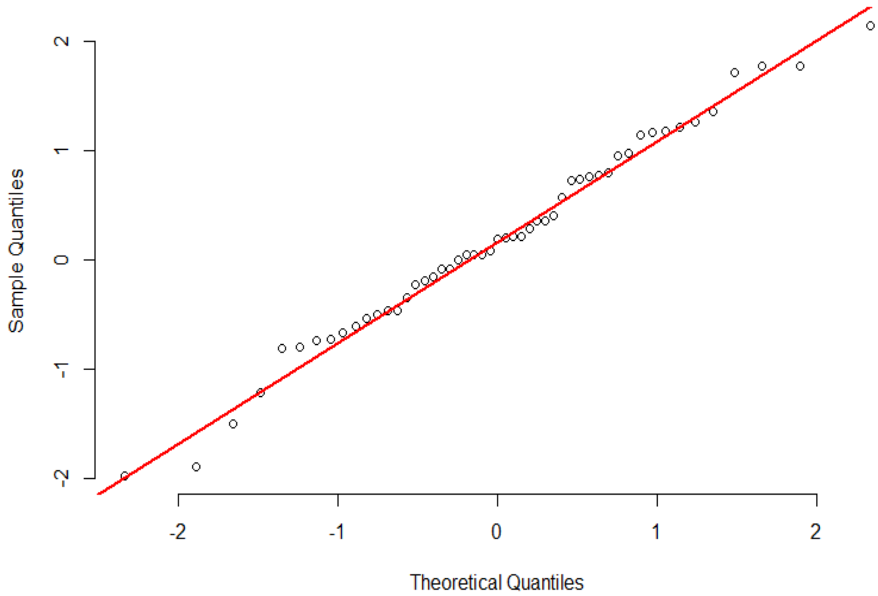

Figure 6.6.11 Normal Q-Q plot of standardized residuals of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2009

Table 6.6.9 Shapiro-Wilk test for normality of the sampling errors of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2009

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.98694	p-value=0.8433

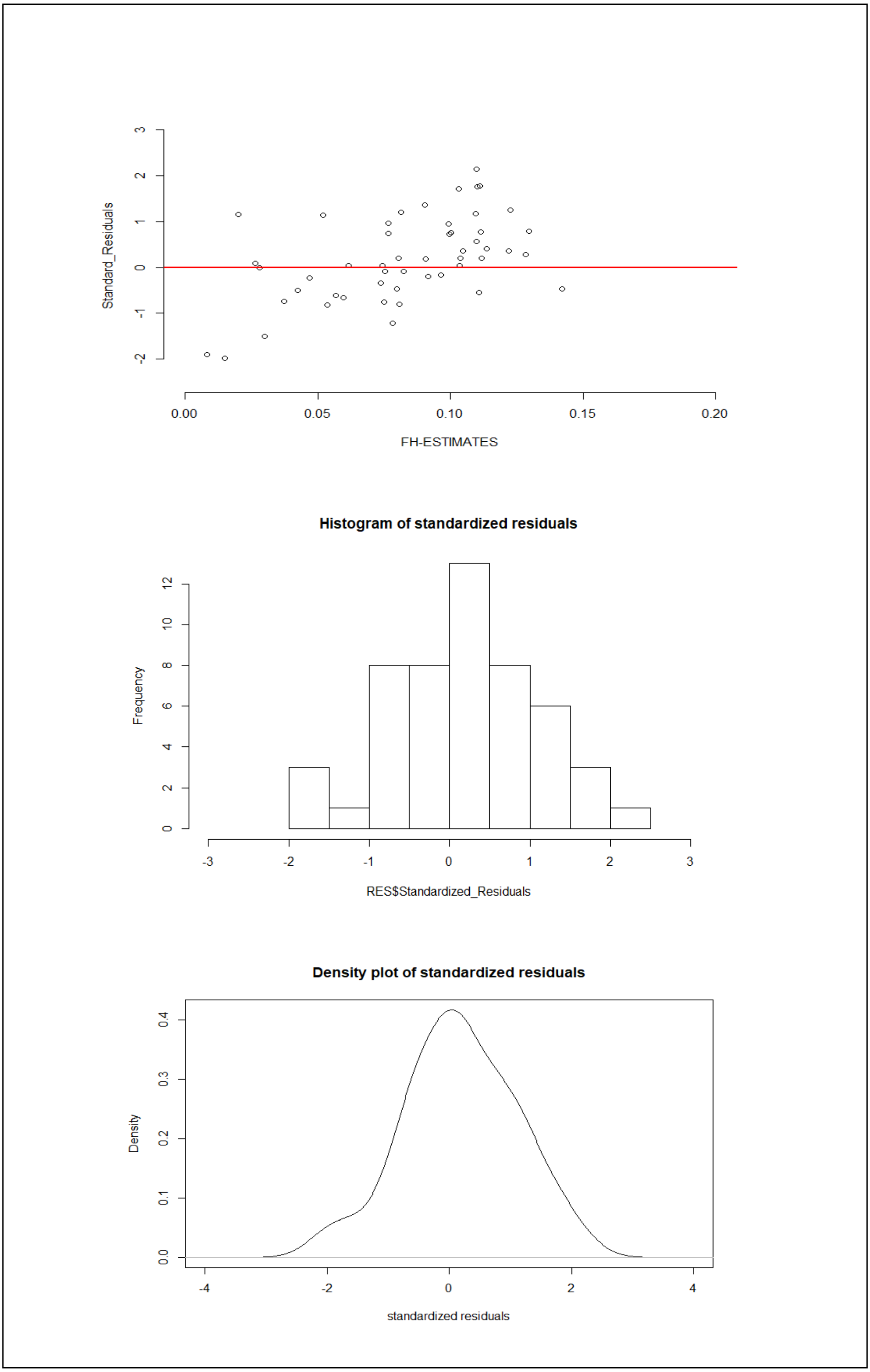

Figure 6.6.12 Residual distribution of the Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2009

iv) A similar procedure as for the sampling errors was followed to check the hypothesis of normality of the random effects. Based on the Q-Q plot (Figure 6.6.13) the random effects lie quite satisfactory on the straight line except for some outliers. The Shapiro-Wilk test was calculated twice. The first time, all prefectures were included in the test and the results are shown in Table 6.6.10. In this case since the p-value is equal to 0.005155 we can reject the null hypothesis of normality of the random effects. But without including in the calculations the two outliers (prefecture of Rodopi and prefecture of Arkadia[87]) the results of the Shapiro-Wilk test changed. In particular, as shown in Table 6.6.11, the p-value is equal to 0.5642 and so in this case we cannot reject the null hypothesis of normality of the random effects. Normality is confirmed also by the histogram and the density plot of random effects (Figure 6.6.14). Furthermore, the cloud of points in the plot of Fay-Herriot model versus random effects has no obvious pattern. In conclusion the hypothesis of normality of random effects seems to be satisfied.

[87] One explanation of these outliers could be the fact that their direct estimates are unexpected small (0.63% for Arcadia and 1.2% for Rodopi) .

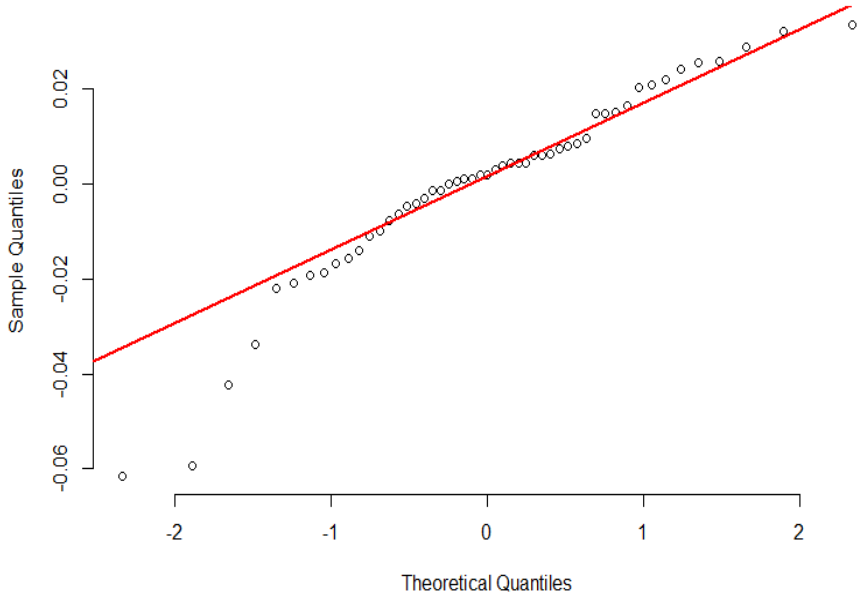

Figure 6.6.13 Normal Q-Q plot of random effects of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2009

Table 6.6.10 Shapiro-Wilk test for normality of the random effects of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2009

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.9304	p-value=0.005165

Table 6.6.11 Shapiro-Wilk test for normality of the random effects of Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2009 (without prefectures of Rodopi (300073) and Arcadia (300012))

Shapiro-Wilk normality test	
data: standardized residuals	
w=0.97997	p-value=0.5642

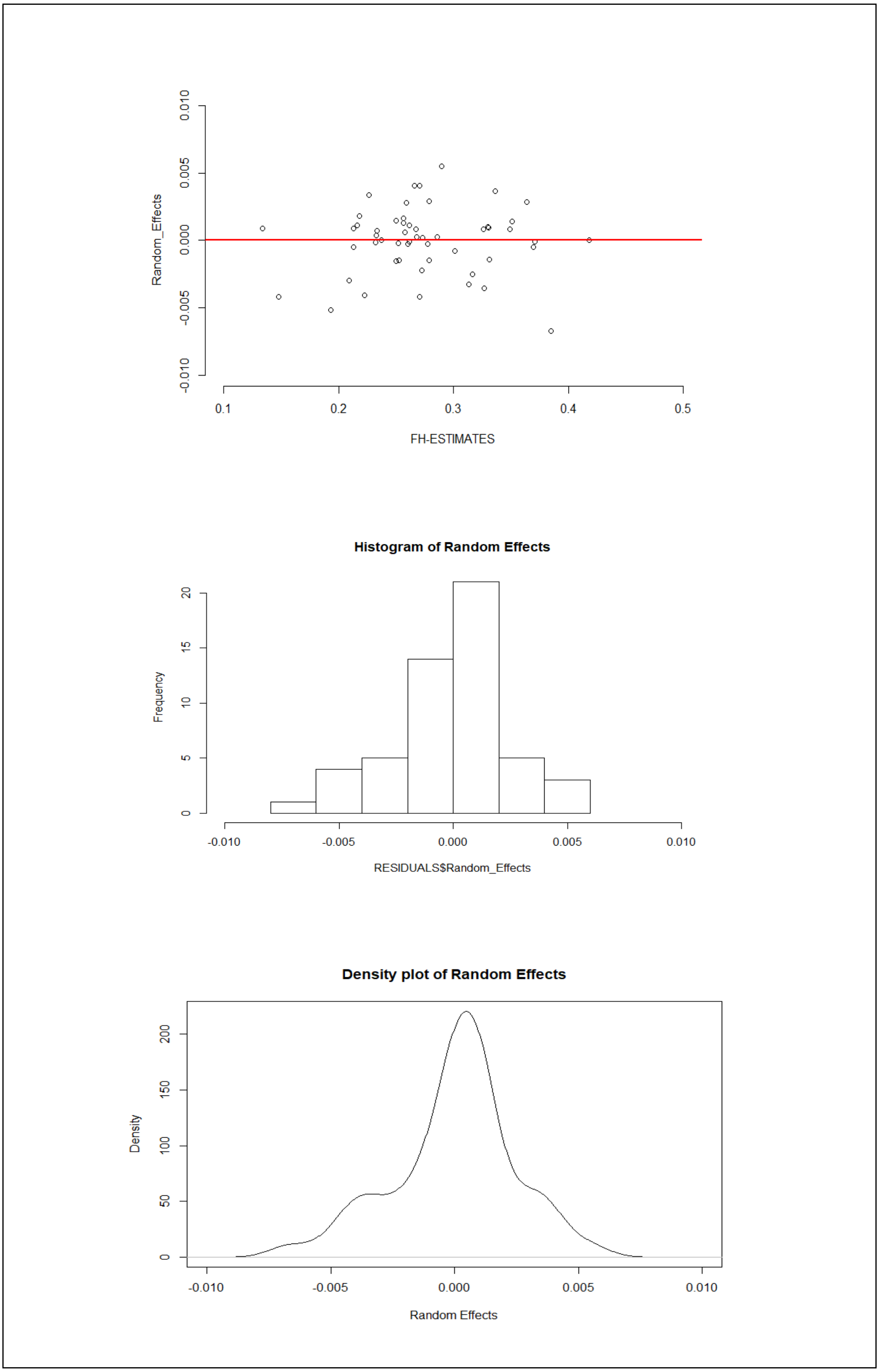

Figure 6.6.14 Random effects distribution of the Fay and Herriot model for the estimation of unemployment rate in Greece for the year 2009

v) In order to evaluate the models, their quality should be checked. As already described (sections 2.6.2 and 7.4 (v and vi)) the coefficient of variation (CV) and the mean square error (MSE) have been used as quality measures. Figures 6.6.15 and 6.6.16 illustrate the standard errors and the CV of direct and Fay-Herriot estimates as well as the precision gain indexes GIP1 and GIP2. Detailed information about direct estimates, Fay and Herriot estimates and the corresponding standard errors, CV, and precision gain indexes GIP1 and GIP2 can be found in the Appendix in Table A20.

Based on Figure 6.6.15 there are three areas (prefecture of Athens, n=1729 code: 300101, prefecture of Serres, n=137, code: 300062 and prefecture of Evrytania, n=6, code: 300005) in which the CV of the direct estimate is less than the CV of the Fay and Herriot estimate (although very close, in the cases of Athens and Serres). In the case of the prefecture of Athens ($CV_{F\text{-}H}$ =9.6284% and CV_{Direct} =9.6265) this is something to be expected since this area has a large enough sample size (n=1729) to give an accurate direct estimate. In the case of the prefecture of Evrytania the sample size is very small (n=6) therefore one would have expected the CV of the Fay-Herriot estimator to be smaller than that of the direct estimator. Nevertheless, there is an overall clear gain of precision when using the Fay-Herriot estimators instead of the direct estimators. This gain is seen in both the standard error ratio (GIP2) and the estimated MSE ratio (GIP1). All the Fay-Herriot estimates have lower standard error than the corresponding direct estimates. Also, the improvement in precision gain tends to be greater for areas with a smaller sample size. Indeed, areas with a small sample size such as the prefectures of Kefallinia (code:300023, n=10, GIP1=3, GIP2=2.44), Kastoria (code:300056, n=39, GIP1=2.16, GIP2=2.30) and Chalkidiki (code:300064, n=42, GIP1=2.25, GIP2=2.30) have a large gain in precision. For example, in Kefallinia the standard error of the Fay-Herriot estimate was reduced 2.16 times and the CV 2.30 times in relation to the direct estimate. This is expected as the direct estimator is likely to be more unstable in areas with small sample size.

Summarizing, the application of small area estimation approaches achieved an overall significant efficiency gain for the estimation of the unemployment rate in Greece for the year 2009.

National statistical offices usually establish a maximum publishable CV. As pointed out by Molina and Marhuenda (2015) and ONS (2004), estimates are considered sufficiently and are suitable for publication when the majority of the CV are below 20%. For the present data the estimated CVs both of the direct and EBLUP F-H estimators of unemployment rate exceeded the level of 20% for 49 (out of the 51).

In conclusion, on the one hand the assumptions of the Fay-Herriot model seem to be satisfied for the unemployment rate in Greece for the year 2009 and on the other there is a clear overall precision gain of the application of the Fay-Herriot model. However, the majority of CV of headcount ratio are over 20%, with only two exceptions.

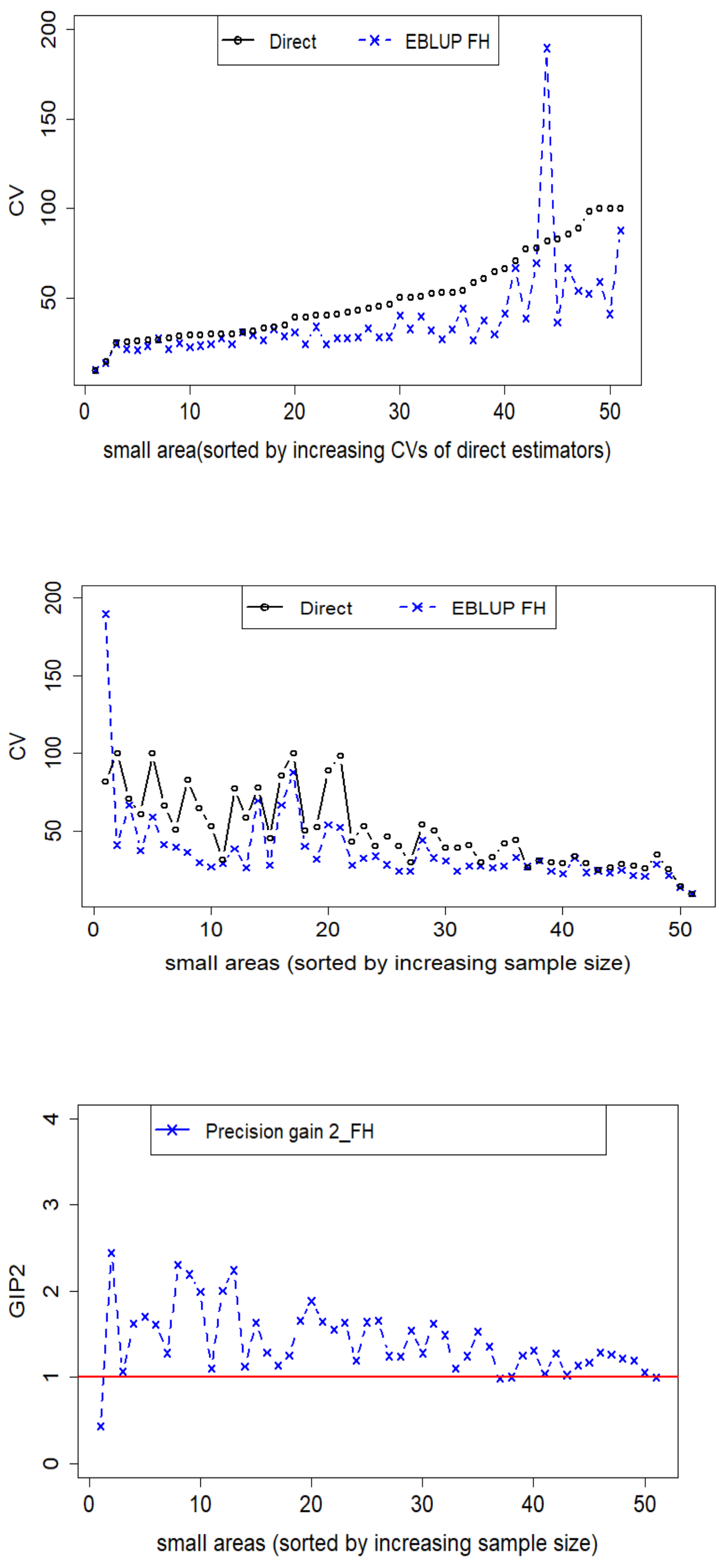

Figure 6.6.15 Coefficients of variation for the Direct and Fay-Herriot estimator of the unemployment rate in Greece for the year 2009 (sorted by increasing sample size and sorted by increasing CVs of direct estimator) and gain in precision index (GIP2)

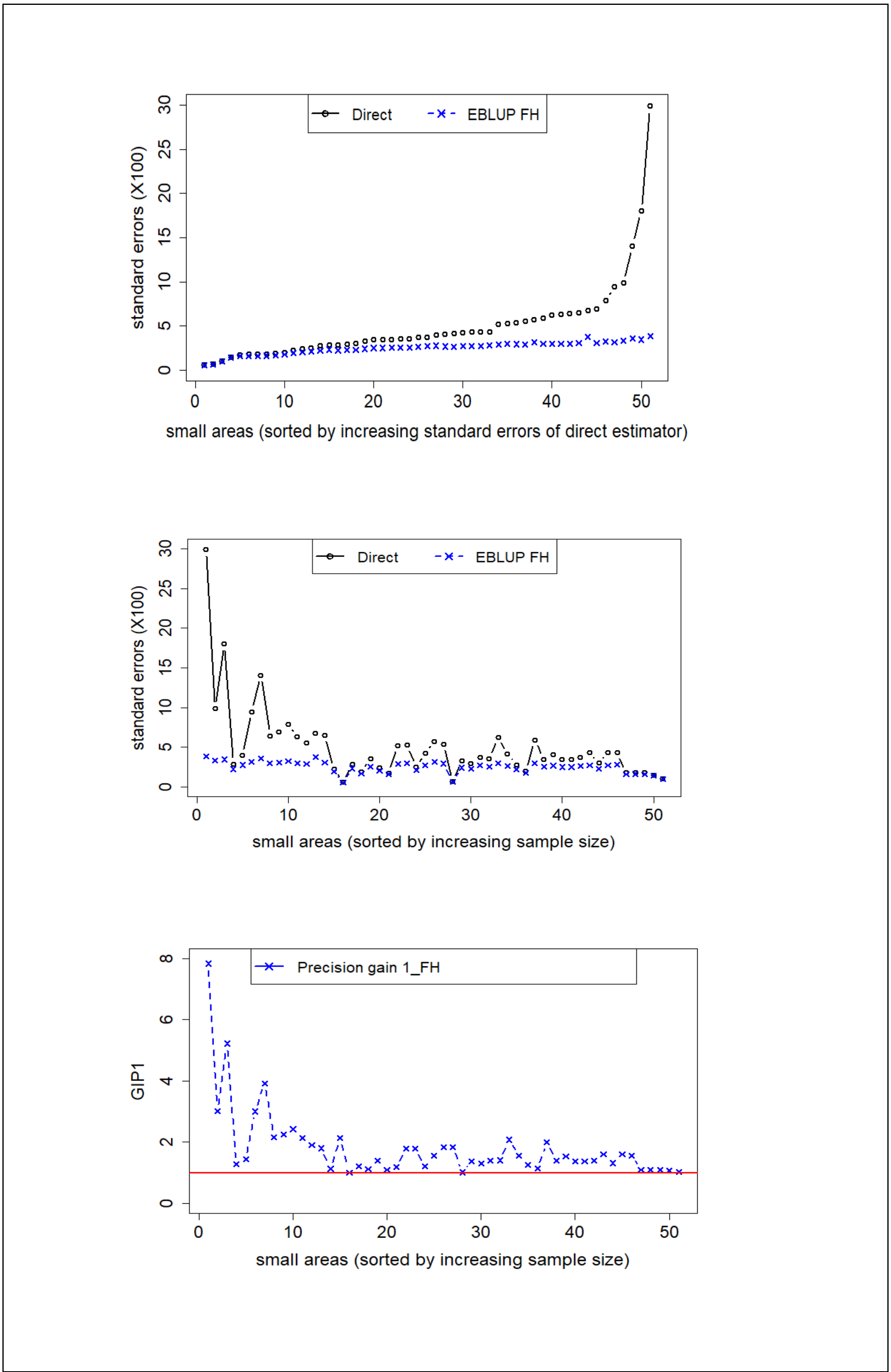

Figure 6.6.16 Standard errors for the Direct and Fay-Herriot estimator of the unemployment rate in Greece for the year 2009 and gain in precision index (GIP1) sorted by increasing sample size

6.7 Results

6.7.1 Results for the estimation of unemployment rate in Greece for the year 2013. The EBLUP Fay-Herriot estimates of the unemployment rate in Greece for the year 2013 are given in detail in Table A17 in the Appendix. Also, direct, and Fay-Herriot estimates of the unemployment rate were placed on the map (Figures 6.7.1-6.7.2) to illustrate the spatial distribution of the analyzed phenomena. Furthermore Figure 6.7.3 show the direct and Fay-Herriot estimates of the unemployment rate in relation to the area specific sample size as well as in relation to the coefficient of variation (CV) of direct estimator.

Concerning the unemployment rate:

- The highest rates occur in the prefectures of Evrytania (41.79%), Pireas (38.47%), Xanthi (37.08%), Fokida (36.96%) and West Attiki (36.36%).
- The lowest rates occur in the prefectures of Lassithi (13.39%), Thesprotia (14.77%), Zakynthos (19.29%) and Evros (20.88%).
- The biggest differences in relation to direct estimates are observed in the prefectures of Argolida (16.52%), Pella (9.82%), Drama (9.41%), Zakynthos (10.84%), Fthiotida (9.69%) and Thesprotia (9.01%). In these areas the sample size is small (n<84).
- In 22 prefectures the difference between direct and Fay-Herriot estimates is over 4%. In these areas there is a reduction (in most areas quite large) for both the MSE and the CV of estimates.

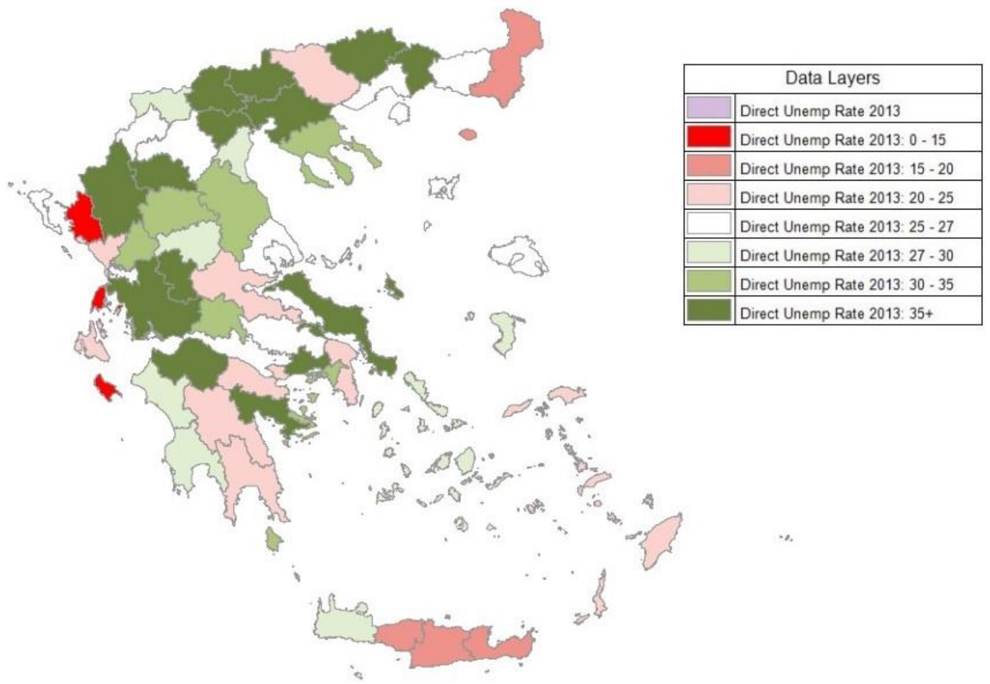

Figure 6.7.1 Cartogram of direct estimates of the unemployment rate in Greece for the year 2013

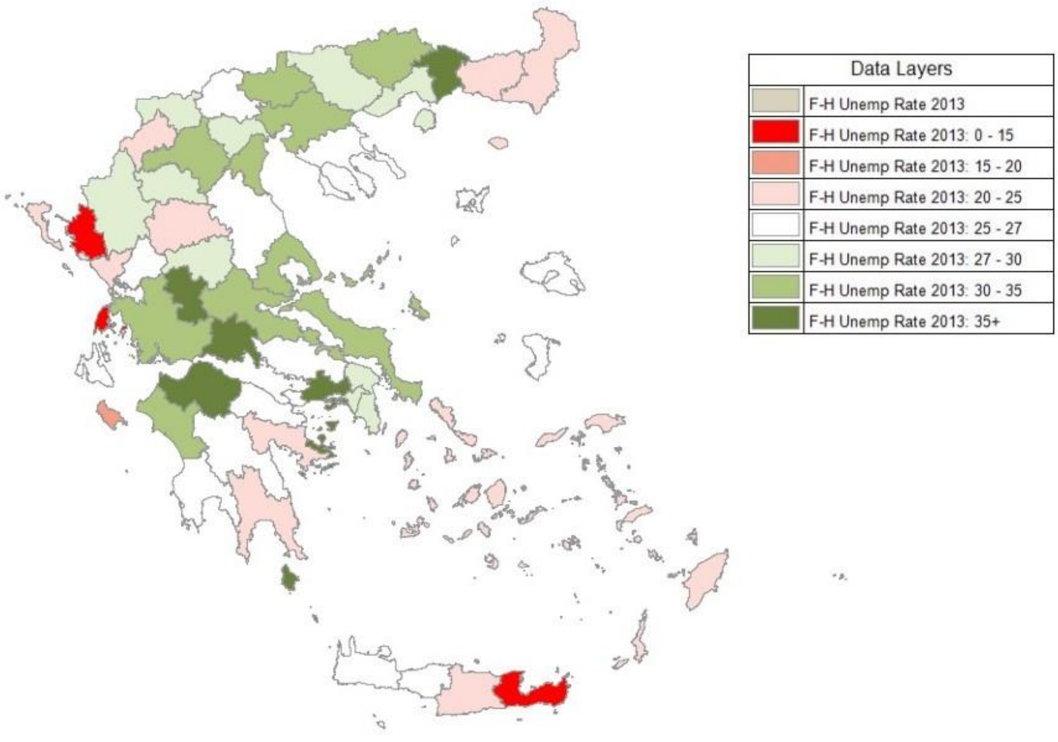

Figure 6.7.2 Cartogram of EBLUP Fay-Herriot estimates of the unemployment rate in Greece for the year 2013

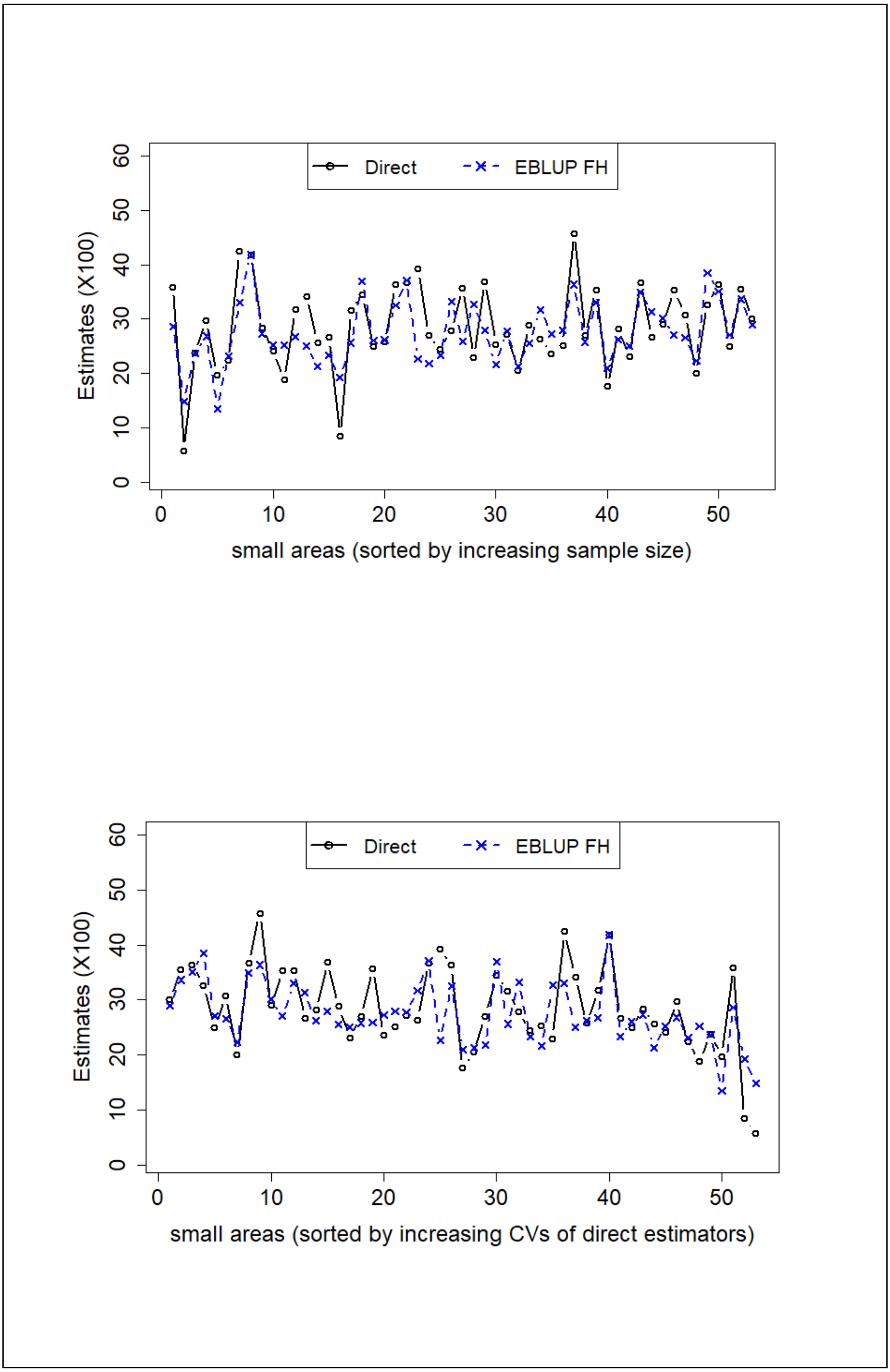

Figure 6.7.3 Direct and Fay-Herriot estimates of the unemployment rate in Greece for the year 2013 sorted by increasing sampling size and by CVs of direct estimators

6.7.2 Results for the estimation of unemployment rate in Greece for the year 2009. The EBLUP Fay-Herriot estimates of the unemployment rate in Greece for the year 2009 are given in detail in Table A18 in the Appendix. Also, direct, and Fay-Herriot estimates of the unemployment rate were placed on the map (Figures 6.7.4 and 6.7.5) to illustrate the spatial distribution of the analyzed phenomena. Furthermore Figure 6.7.6 shows the direct and Fay-Herriot estimates of the unemployment rate in relation to the area specific sample size as well as in relation to the coefficient of variation (CV) of direct estimator.

Concerning the unemployment rate:

- The highest unemployment rates occur in the prefectures of Chios (14.20%), Achaia (12.97%), Kozani (12.83%), Evia (12.25%) and Xanthi (12.20%).
- The lowest unemployment rates occur in the prefectures of Arkadia (0.81%), Rodopi (1.5%), Evrytania (2.01 %), Lassithi (2.65%) and Lakonia (2.16%).
- The biggest differences in relation to direct estimates are observed in the prefectures of Zakynthos (20.3%), Fokida (18.6%), Serres (11.1%), Kilkis (9.81%) and Preveza (9.69%). From these areas, the prefectures of Zakynthos, Fokida and Preveza have a small sample size (n<48).
- In 12 prefectures the difference between direct and Fay-Herriot estimates is over 3%. In these areas there is a reduction (in most areas quite large) for both the MSE and the CV of estimates (the prefectures of Evrytania and Serres are an exception).

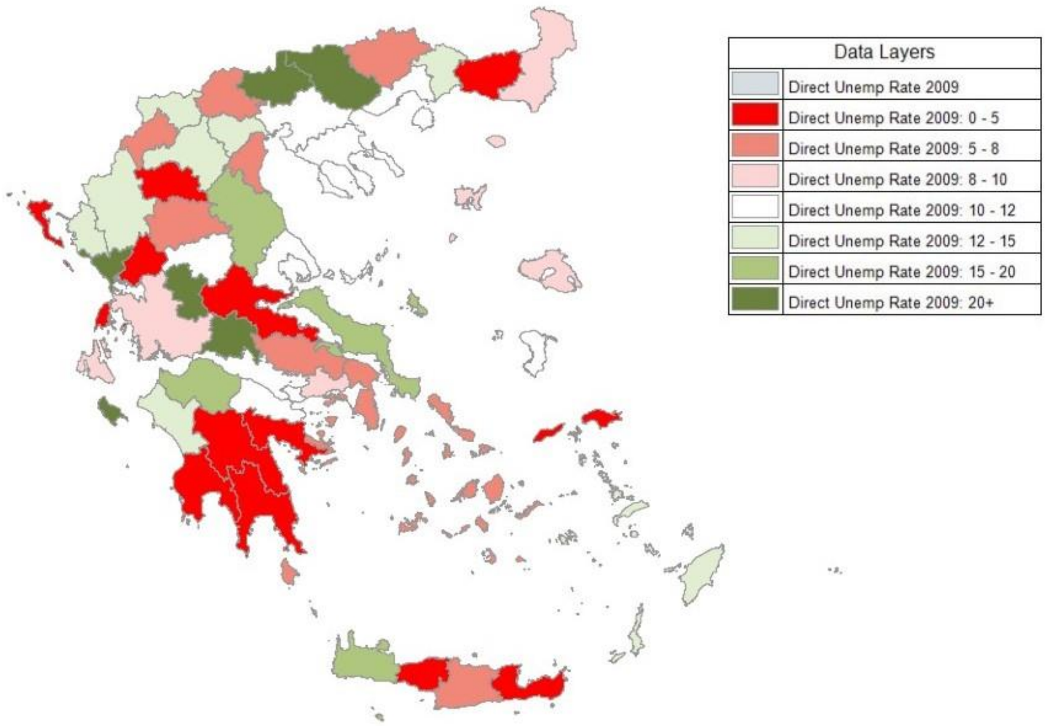

Figure 6.7.4 Cartogram of direct estimates of the unemployment rate in Greece for the year 2009

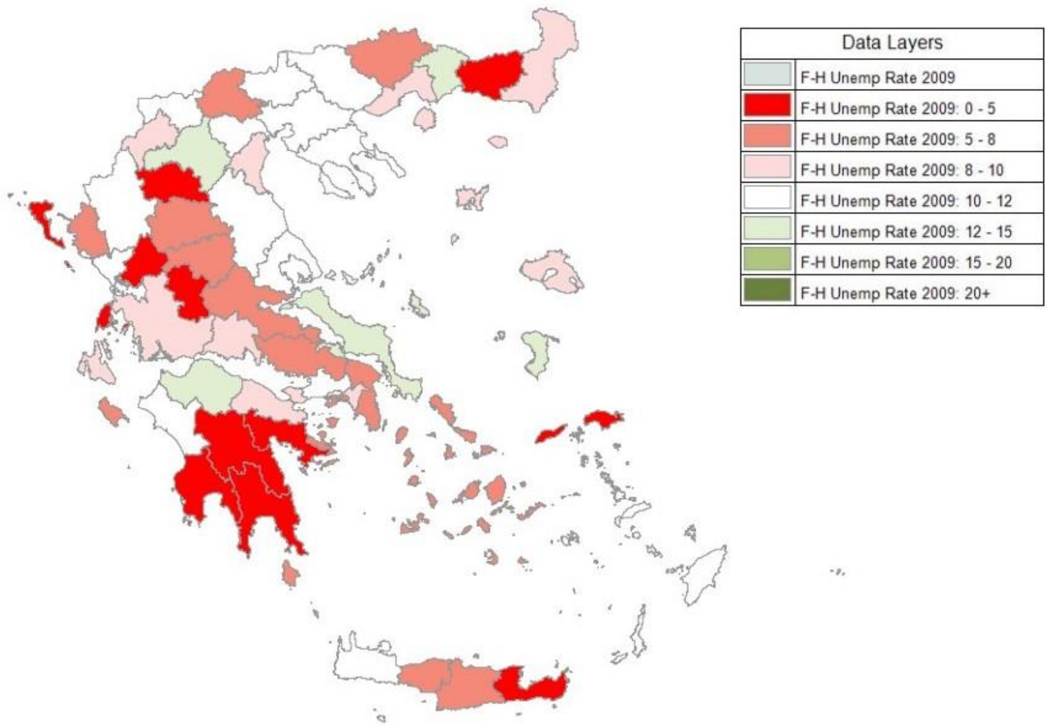

Figure 6.7.5 Cartogram of EBLUP Fay-Herriot estimates of the unemployment rate in Greece for the year 2009

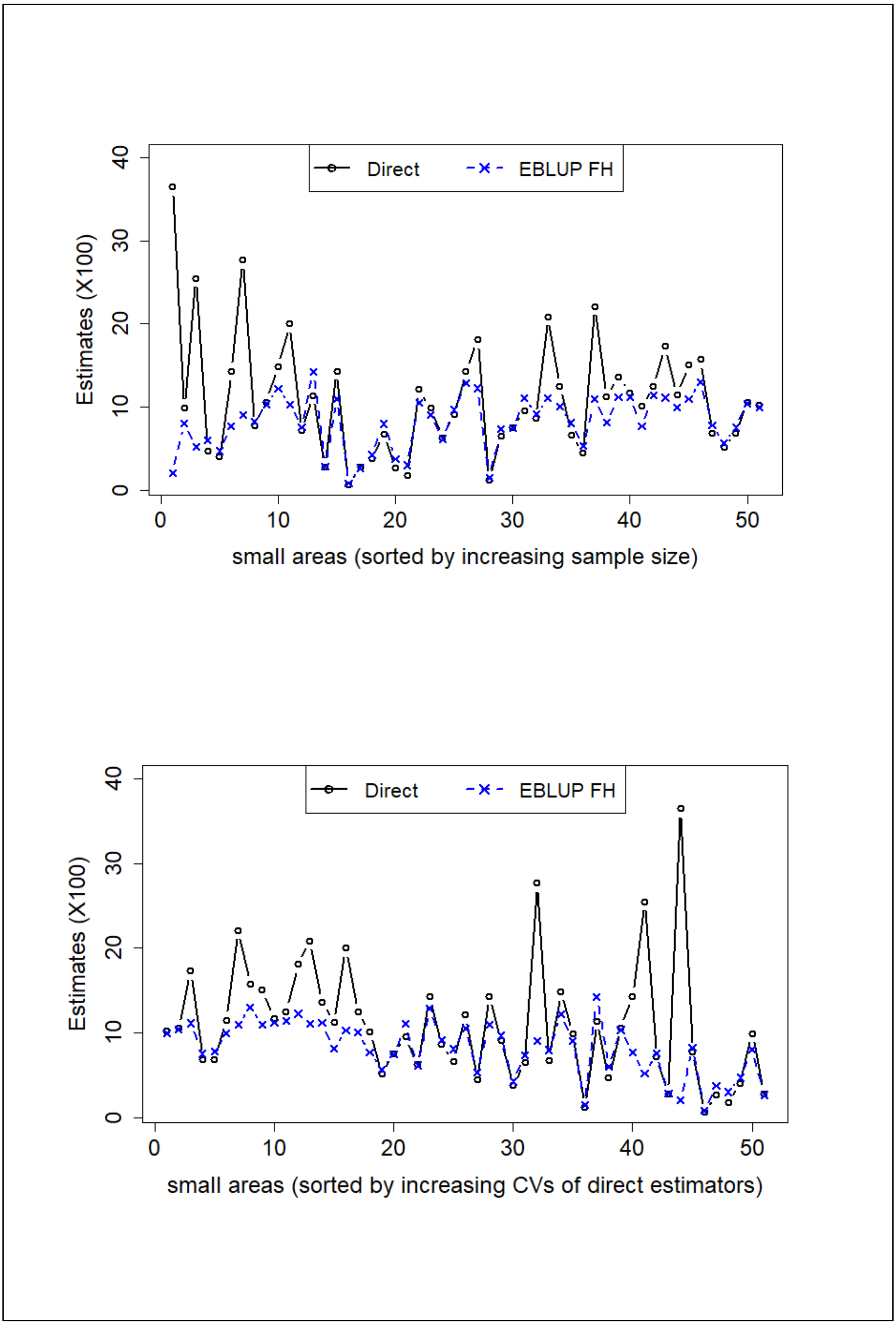

Figure 6.7.6 Direct and Fay-Herriot estimates of the unemployment rate in Greece for the year 2009 sorted by increasing sampling size and by CVs of direct estimators

6.7.3 Comparative results for the estimation of unemployment rate in Greece for the years 2009 and 2013. The EBLUP Fay-Herriot estimates of the unemployment rate in Greece for the years 2009 and 2013 are given in detail in Table A21 in the Appendix. Also, Figures 6.7.7 and 6.7.9 show the EBLUP Fay-Herriot estimates of the unemployment rate for the year 2013 compared to the year 2009. Furthermore, a scatter plot was created for the unemployment rate estimates of 2009 versus 2013 to examine whether there is a pattern between the two years of data (Figure 6.7.8).

Concerning the estimation results:

- There was an increase in unemployment rates in all the prefectures of Greece in the year 2013 compared to 2009.
- The biggest differences occurred in the prefectures of Pireas (30.65%), Fokida (27.9%), Fthiotida (27.3%), Etolia and Akarnania (25.7%). In 24 prefectures the increase is over 20%. In the prefectures of Etolia and Akarnania, Pieria, Achaia, Thessaloniki and West Attiki there was a large increase in both unemployment and poverty.
- The smallest differences occurred in the prefectures of Thesprotia (7.9%), Dodekanissos (9.9%) and Lassithi (10.74%). In these areas there were also small differences in poverty, both in the headcount ratio and the poverty gap rates.

Also, based on the scatter plot (Figure 6.7.9) there is no trend to the data.

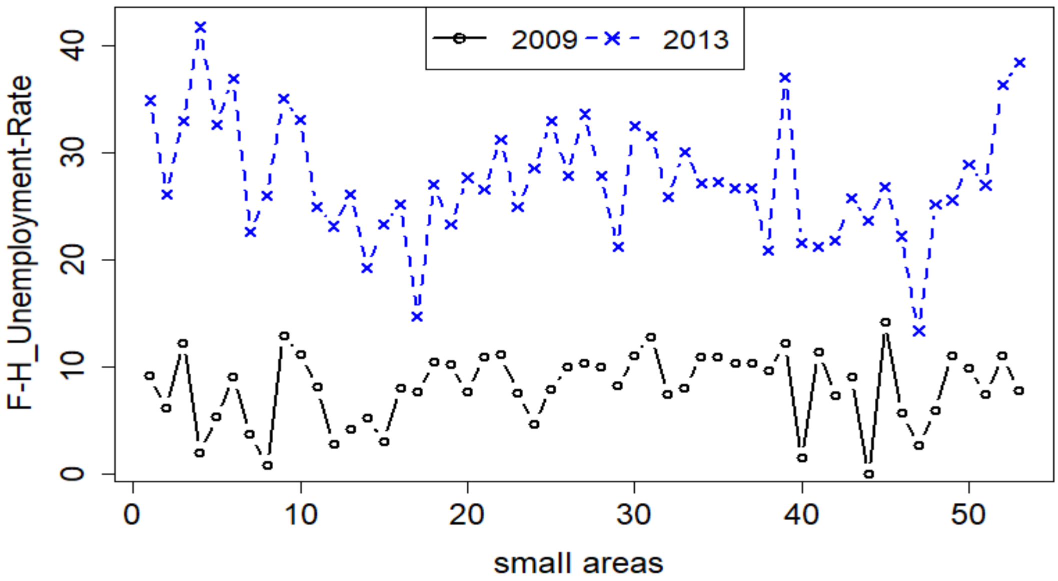

Figure 6.7.7 EBLUP F-H estimates of the unemployment rate in Greece for the years 2009 and 2013

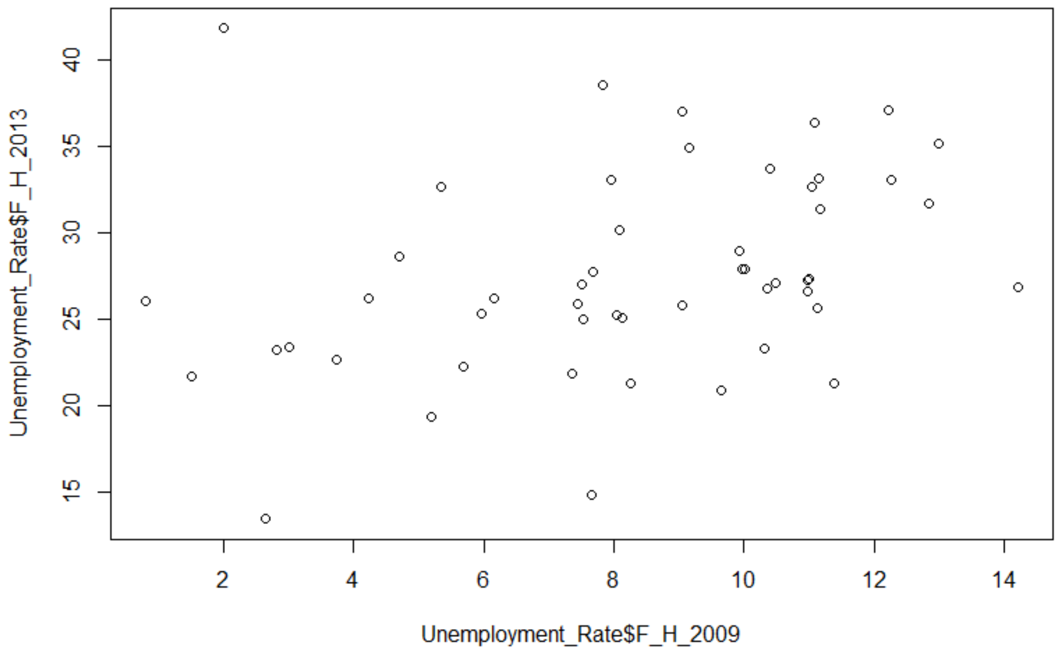

Figure 6.7.8 Scatter plot for the unemployment rate in 2009 versus the unemployment rate in 2013

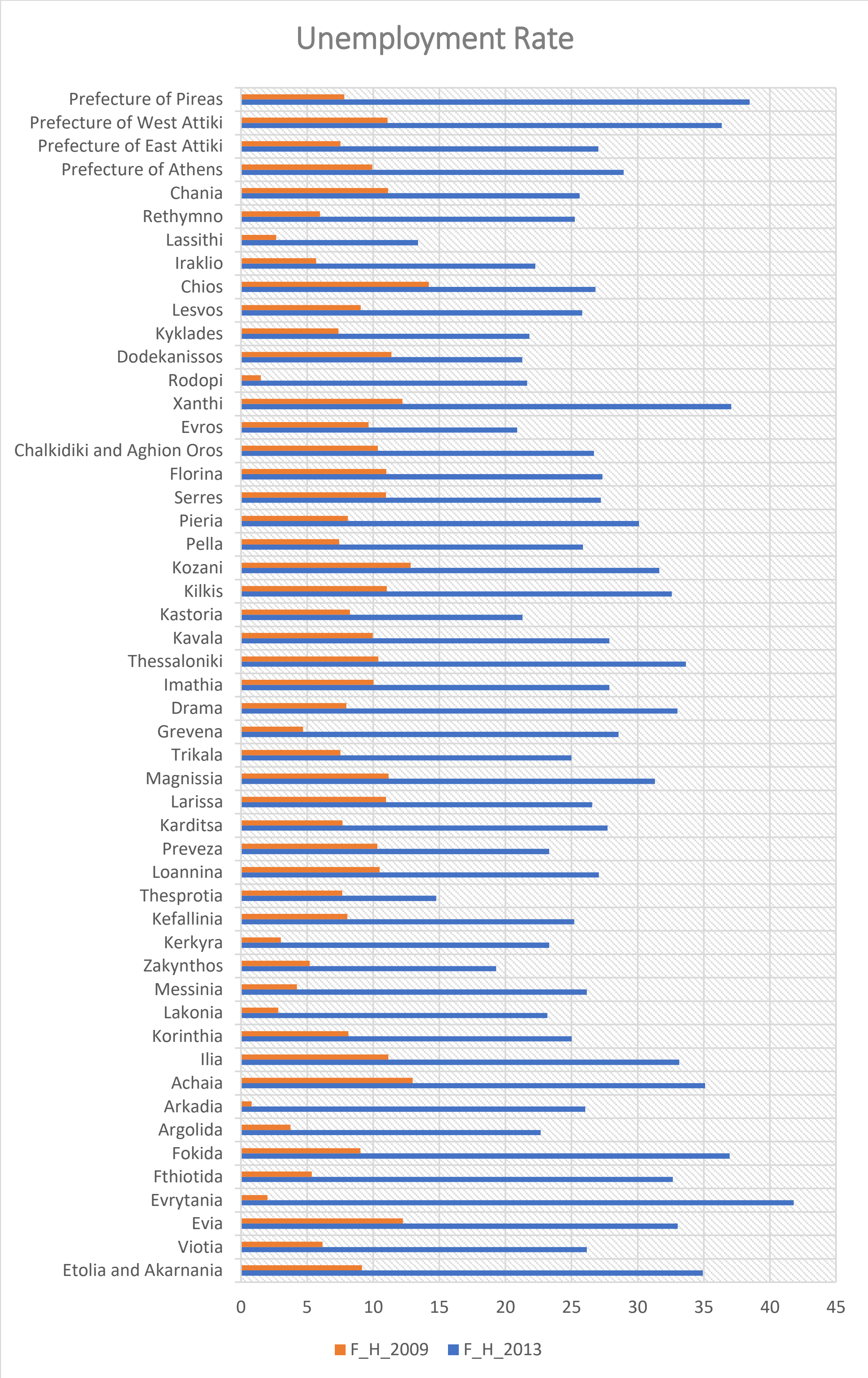

Figure 6.7.9 Comparative bar chart for the EBLUP F-H estimates (%) of the unemployment rate in Greece for the years 2009 and 2013

7. Conclusions and suggestions for further work

In this thesis SAE methods were used to estimate poverty and unemployment in Greece at the level of sub regions-NUTS 3 (*Nomoi*). The EBLUP estimator based on the Fay and Herriot model was adopted to provide estimates of the headcount ratio, the poverty gap and the unemployment rate of the Greek population at two different times, in 2009 (shortly before the start of the Greek financial crisis) and 2013 (during the crisis). The above work was carried out combining survey data from the EU-SILC of 2009 and 2013 with auxiliary data derived from the 2001 and 2011 Census, respectively.

Fay and Herriot model consists of a sampling model for the direct estimates and a linking model for the parameters of interest. For the sampling model, the Horvitz-Thompson direct estimates of the target parameters were derived for each small area i (*Nomoi)* for the years 2009 and 2013 using the unit level data from the EU-SILC surveys of 2009 and 2013 respectively. For the linking model, area-specific auxiliary data from the national Greek Census of 2001 and 2011 were used. A total of 19 auxiliary variables from the 2001 Census and 32 auxiliary variables from the 2011 Census were examined. The initial set of auxiliary variables was selected based on the factors that seem to influence poverty and unemployment the most, according to the literature as well as based on past researches in estimating poverty and unemployment using SAE methods. The availability of Census data also played an important role in the selection of the variables.

In order to build the optimal small area model a three-phase variable selection process was performed. The final model for each target parameter and each year was chosen according to three information criteria (AIC, BIC, cAIC). Also, the covariates included in the final models were significant (p-values of the significance of each coefficient less than 10%) and the sign of the estimated model coefficients (beta parameter) was justified according to knowledge of the analyzed phenomena from the literature.

For the year 2013, the model chosen for the estimation of headcount ratio and poverty gap contained three auxiliary variables: percentage of people per prefecture with low education (X1), percentage of people per prefecture aged under 15 years old (X6) and percentage of women per prefecture (X10). The model chosen for the estimation of unemployment rate contained five auxiliary variables: percentage of people per prefecture aged 29-49 years old (X7), percentage of people per prefecture aged over 50 years old (X8), percentage of inactive people per prefecture (X13), percentage of nuclear families per prefecture without an employed member (X30), percentage of single-parent families per prefecture (X31).

For the year 2009, the model chosen for the estimation of headcount ratio and poverty gap contained two auxiliary variables: percentage of people per prefecture with low education (X1) and percentage of inactive people per prefecture (X13). The model chosen for the estimation of unemployment rate contained two auxiliary variables: percentage of people per prefecture over 65 years old (X9) and percentage of inactive people per prefecture (X13).

Small area estimation approaches depend to a large extent on the quality of the models used. Wrong specifications can lead to a strong bias of the estimators and accordingly to a misleading information base in the applications. Therefore, after applying small-area estimation, the model chosen must be carefully examined and checked for a violation of the underlying assumptions and a possible bias (Rao and Wu, 2001). Different diagnostic tools were applied to assess the fit and the performance of the small area estimators as well as to check the reliability of the results.

The selected models were carefully examined and checked for a violation of the underlying assumptions and a possible bias. As far as the year 2013 is concerned, the assumptions of the F-H model seem to be satisfied for all parameters of interest, that is headcount ratio, poverty gap and unemployment rate. Concerning the year 2009, the assumptions of the F-H model seem to be satisfied for the headcount ratio and unemployment rate while in the case of poverty gap the only assumption that is not clearly satisfied is that of the constant variance of the sampling errors.

In order to evaluate the models, their quality should be checked. The coefficient of variation (CV) and mean square error (MSE) were used as quality measures for the final selected models. The reduction in both the MSE and the CV of the estimates was clear in most cases. The only domain where the CV of the F-H estimator was greater than the CV of the direct estimator was the prefecture of Athens for the estimation of unemployment rate the year 2013. Something to be expected (Molina and Rao, 2010) since this area has a large enough sample size so it can give an accurate direct estimator. So, there was an overall significant efficiency gain for estimating poverty and unemployment using the EBLUP F-H instead of direct estimator. The improvement in precision gain tended to be greater for areas with a smaller sample size. This is expected as the direct estimator is likely to be more unstable in areas with small sample size (Molina and Morales, 2009). Therefore, the accuracy of the estimators of poverty and unemployment was increased using the small area methods.

National statistical offices usually establish a maximum publishable CV. As pointed out by Molina and Marhuenda (2015) and ONS (2004), estimates are considered precise and are suitable for publication when the majority of CV are below 20%. In the present study, for the year 2013 the majority of the CV of the EBLUP F-H estimators were below 20%. Specifically:

- In the case of headcount ratio, the estimated CVs of the EBLUP F-H estimators exceeded the level of 20% for only three domains, compared to 13 domains for the direct estimators.
- In the case of the poverty gap, the estimated CVs of the EBLUP F-H estimators exceeded the level of 20% for 9 domains, compared to 22 domains for the direct estimators.
- In the case of the unemployment rate, the estimated CVs of the EBLUP F-H estimators exceeded the level of 20% for only one domain, compared to 34 domains for the direct estimators.

For the year 2009:

- In the case of the headcount ratio, the estimated CVs of the EBLUP F-H estimators exceeded the level of 20% for seven domains, compared to 19 domains for the direct estimators.
- In the case of the poverty gap, the estimated CVs of the EBLUP F-H estimators exceeded the level of 20% for 19 domains, compared to 30 domains for the direct estimators.
- In the case of the unemployment rate, the estimated CVs of the EBLUP F-H estimators (as well as the direct estimators) exceeded the level of 20% for most of the domains (49 out of 51).

Based on the above results, as far as the year 2013 is concerned, the EBLUP F-H estimators for all the target parameters are considered precise and are suitable for publication. As far as the year 2009 is concerned, the EBLUP F-H estimators of the headcount ratio are considered precise and are suitable for publication while the same does not seem to be valid for the poverty gap and especially for the unemployment rate.

This study has both strengths and limitations. The first limitation was the availability of both auxiliary and survey data. The only available source for the auxiliary data was the Greek Census. Census data are the best source as their area-based values are free of sampling errors. However, Censuses are conducted only once a decade. Thus, for the estimation of the target parameters of 2009, the 2001 Census was selected, although there is a considerable time gap. Also, the content of the 2001 Census was quite restricted compared to that of 2011. As a result, the variables extracted and used in the 2001 Census were almost half compared to those of the 2011 Census. Based on the research results, these two factors (time gap and few variables) could be an explanation for the lower performance of the 2009 small area models. As pointed out by Rao (2003) and Whitworth (2013), the availability of good auxiliary data is crucial to the formulation of small area models. The success of any model-based method depends on how good the auxiliary data are as predictors of the study variables. Furthermore, Census data was available only at area level for confidentiality reasons. Therefore, among the various small area models, the F-H model was chosen as it can be applied when area level auxiliary data is available.

The available source for the sample data was the EU-SILC survey in Greece. The EU-SILC data, for both 2009 and 2013, were available at unit level upon request to ELSTAT. EU-SILC data were used to estimate both poverty and unemployment. As far as unemployment is concerned the initial goal was the use of microdata from the Labor Force Survey (LFS) in Greece for the years 2009 and 2013. However, for reasons of confidentiality and protection of personal data, the microdata was not provided. Thus, the data of EU-SILC that had already been provided were used. Nevertheless, it should be noted that there are some minor differences in the definition of unemployed persons between EU-SILC and the Census.

Furthermore, there were limitations during the estimation process. Some prefectures dropped out of the estimation process as the numerical value of the direct estimate of the target parameters for the specific prefectures was zero. For the year 2013, this applied to the prefecture of Lefkada for the estimation of unemployment rate. For the year 2009, it applied to the prefectures of Lefkada, Arta and Samos for the estimation of unemployment rate and the prefecture of Kefalonia for the estimation of headcount ratio and poverty gap.

Concerning the estimation results for the year 2013: in 10 prefectures the difference between direct and Fay-Herriot estimates of the headcount ratio was over

3%, while in 16 prefectures the difference was over 2%. As far as the poverty gap is concerned, in eight prefectures the difference was over 2%, while in six prefectures the difference was over 4%. Concerning the unemployment rate, in 22 prefectures the difference between direct and Fay-Herriot estimates was over 4%. In all these areas there was a reduction (in most areas quite large) in both the MSE and the CV of estimates.

Concerning the estimation results for the year 2009: in 15 prefectures the difference between direct and Fay-Herriot estimates of the headcount ratio was over 3%. As far as the poverty gap is concerned, in seven prefectures the difference was over 2%, while in six prefectures the difference was over 3%. Concerning the unemployment rate, in 12 prefectures the difference between direct and Fay-Herriot estimates was over 3%. In all these areas there was a reduction (in most areas quite large) for both the MSE and the CV of estimates.

Comparing the results of the two years there was a large increase in unemployment rate in 2013 compared to 2009 in all the prefectures of Greece. In 24 prefectures the increase was over 20%. The large increase in unemployment is justified by the financial crisis that took place in Greece by the end of 2009. As far as poverty rate is concerned, in 30 prefectures there was a reduction for the year 2013 compared to 2009. This is an unexpected result as in 2013 one would expect an increase due to the financial crisis. However, in 36 prefectures there was an increase in the poverty gap rate for the year 2013 compared to 2009. Furthermore, in areas with a large increase in the unemployment rate such as Etolia and Akarnania, Pieria, Achaia, Thessaloniki and West Attiki there was also a large increase in both headcount ratio and poverty gap. On the other hand, in areas with a small increase in the unemployment rate such as Thesprotia, Dodekanissos and Lassithi there was a small change in both headcount ratio and poverty gap.

Despite the limitations and weaknesses, the present study can contribute to the constantly growing research on the application of SAE methods in the estimation of social characteristics in small geographical areas. The adequate knowledge of poverty and unemployment in small areas allows governments to develop appropriate policies and programs and target them effectively.

For further research, the author plans to apply SAE methods to other social characteristics besides poverty and unemployment, such as social exclusion. In addition to the F-H model, it is planned to implement other small area models and compare their performance.

References

Ελληνόγλωσση

Ν. 3852/2010 «Νέα Αρχιτεκτονική της Αυτοδιοίκησης και της Αποκεντρωμένης Διοίκησης-Πρόγραμμα Καλλικράτης», (ΦΕΚ 87/τ.Α΄/07-06-2010). Εφημερίς της Κυβερνήσεως.

Ξενόγλωσση

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(3), 370-374.

Ambler, R., Caplan, D., Chambers, R., Kovacevic, M. & Wang, S. (2001). Combining unemployment benefits data and LFS data to estimate ILO unemployment for small areas: An application of a modified Fay-Herriot method. In: *Proceedings of the International Association of Survey Statisticians*, Meeting of the International Statistical Institute: Seoul, August 2001.

Atkinson, A. B., Cantillon, B., Marlier, E., & Nolan, B. (2002). *Social indicators: The EU and social inclusion.* Oxford: Oxford University Press.

Bartholomew, D., Moore, P., Smith, F., & Allin, P. (1995). The Measurement of unemployment in the UK. *Journal of the Royal Statistical Society, Series A, 158*(3), 363-417.

Battese, G.E., Harter, R.M., & Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data, *Journal of the American Statistical Association, 83,* 28-36.

Bell, W.R. (1999). Accounting for uncertainty about variances in small area estimation. *Bulletin of the International Statistical Institute*, 52nd Session, Helsinki. Retrieved from http://www.census.gov/did/www/saipe/publications/files/Bell99.pdf.

Bell, W.R. (2009). The U.S. Census Bureau's small area income and poverty estimates program: A statistical review. Retrieved from http://cio.umh.es/data2/T1A%20William.R.Bell@census.gov.pdf.

Biagi, F., & Bocconi, L.C. (2005). Demographic and education effects on unemployment in Europe: Economic factors and labour market institutions. *IZA Discussion Paper* 1806, Institute for the Study of Labor (IZA).

Brandolini, A., & Viviano, E. (2018). Measuring employment and unemployment. *IZA World of Labor*, Institute of Labor Economics (IZA), 445, August 2018. doi: 10.15185/izawol.445.

Brooks, S.P. (1988). Markov chain Monte Carlo method and its applications. *Statistician*, *47,* 69-100.

Brown, G., Chambers, R., Heady, P., & Heasman, D. (2001). Evaluation of small area estimation methods: An application to unemployment estimates from the UK LFS. *Proceedings of Statistics Canada Symposium Achieving Data Quality in a Statistical Agency: A Methodological Perspective,* Statistics Canada. doi:10.12 01/9780203166314.ch1.

Castleman, T., Foster J., & Smith S. (2015). Person equivalent headcount measures of poverty. *Institute for International Economic Policy Working Paper, Series No. IIEP-WP-2015-10.* Washington, D.C.: The George Washington University. Retrieved from https://www2.gwu.edu/~iiep/assets/docs/papers/2015WP/Cast lemanFosterSmithIIEPWP201510.pdf.

Chandra, H., & Chambers, R. (2007). Small area estimation for skewed data. *Small Area Estimation Conference*, Pisa (Italy).

Chambers, R., & Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, *93*, 255–268.

Claeskens, G., & Hjort, N.L. (2008). *Model selection and model averaging*. Cambridge University Press.

Clemenceau, A., & Museux, J.-M. (2007). EU-SILC (community statistics on income and living conditions: general presentation of the instrument). In Eurostat (Ed.), *Comparative EU statistics on income and living conditions: Issues and challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006),* Luxembourg: Office for Official Publications of the European Communities, pp. 11-36. Retrieved from http://epp.eurostat.ec.europa.eu/cache /ITY_OFFPUB/KS-RA-07-007/EN/ KS-RA-07-007-EN.PDF.

Cochran, W.G. (1977). *Sampling techniques.* 3rd Edition, New York: Wiley

Council Decision 75/458/EEC (1975, 22 July). *Concerning a programme of pilot schemes and studies to combat poverty.* 75/458/EEC (OLJ 99/3430.7.75.).

Cressie, N. (1992). REML estimation in Empirical Bayes smoothing of Census undercount. *Survey Methodology, 18,* 75-94.

D'Aló, M., Consiglio, L.D., Falorsi, S., Solari, F., Pratesi, M., Salvati, N., & Ranalli, M. (2012). Small area models for unemployment rate estimation at sub-provincial areas in Italy. *Journal of the Indian Society of Agricultural Statistics, 66*(1), 43-53.

Dalton, H. (1920). The measurement of the inequality of incomes. *Economics Journal, 30(119),* 348-361.

Dandekar, V.M., & Rath, N. (1971). *Poverty in India.* Indian School of Political Economy, Poona.

Datta, G.S. (2009). Model-based approach to small area estimation. In *Sample Surveys: Inference and Analysis,* (D. Pfeffermann and C. R. Rao, Eds.). *Handbook of Statistics, 29B,* pp. 251–288. North-Holland, Amsterdam.

Datta, G.S., & Ghosh, M. (1991). Bayesian prediction in linear models: Application to small area estimation. *The Annals of Statistics, 19*(4), 1748-1770.

Datta, G.S., Ghosh, M., Nangia, N., & Natarajan, K. (1996). Estimation of median income of four-person families: A Bayesian approach. *Bayesian Analysis in Statistics and Econometrics,* 129-140, New York: Wiley.

Datta, G.S., Day, B., & Basawa, I. (1999). Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation. *Journal of Statistical Planning and Inference, 75,* 269-279.

Decancq, K., Goedemé, T., Van den Bosch, K., & Vanhille, J. (2013). The Evolution of poverty in the European Union: Concepts, measurement and data. *ImPRovE Methodological Paper* 13/01. Antwerp: Herman Deleeck Centre for Social Policy, University of Antwerp.

Deely, J.J., & Lindley, D.V. (1981). Bayes empirical Bayes. *Journal of the American Statistical Association, 76,* 833-841.

Deville, J.C., & Sarndal, C.E. (1992). Calibration estimation in survey sampling. *Journal of the American Statistical Association, 87,* 376-382.

Dick, J. P. (1995). Modelling net undercoverage in the 1991 Canadian Census. *Survey Methodology, 21,* 44–55.

Dickerson, A., & Popli G. (2014). Persistent poverty and children's cognitive development: Evidence from the UK millennium cohort study. *Sheffield*

Economic Research Paper. Series No. 2011023. Sheffield: University of Sheffield.

EEC. (1985). *On specific community action to combat poverty* (Council Decision of 19 December 1984), 85/8/EEC. *Official Journal of the EEC, 2,* 24-25.

Elazar, D.N. (2004). Small area estimation of disability in Australia. *Statistics in Transition,* 6(5), 667-684.

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

ESSnet Project on Small Area Estimation (2012a). *Report on workpackage 3: Quality assessment,* Final Version, March 2012. Retrieved from https://ec.europa.eu/eurostat/cros/content/sae-finished_en.

ESSnet Project on Small Area Estimation (2012b). *Report on workpackage 4: Software tools,* Final Version, March 2012. Retrieved from https://ec.europa.eu /Eurostat/cros/system/files/WP4report_0.pdf.

ESSnet Project on Small Area Estimation (2012c). *Report on workpackage 6: Guidelines,* Final Version, March 2012. Retrieved from https://ec.europa.eu/eurostat/cros/system/files/WP6-Report.pdf.

ESSnet Project on Small Area Estimation (2012d). *Report on workpackage 5: Case studies,* Final Version, March 2012. Retrieved from https://ec.europa.eu/eurostat/cros/system/files/ESSnet%20SAE%20WP5%20Report-final-rev2.pdf.

ESSnet Project on Small Area Estimation (2014). *Project on handbook on methodology of modern business statistics,* (Memobust Handbook), Module on Synthetic Estimators for SAE. Retrieved from https://ec.europa.eu/eurostat/cros /content/synthetic-estimators-small-area-estimation-method_en.

ESSnet Project on Small Area Estimation (2014b). *Project on handbook on methodology of modern business statistics,* (Memobust Handbook), Module on GREG Estimators for SAE. Retrieved from https://ec.europa.eu/eurostat/cros/content/generalized-regression-estimator-method_en.

ESSnet Project on Small Area Estimation (2014c). *Project on handbook on methodology of modern business statistics,* (Memobust Handbook), Module on Composite Estimators for SAE. Retrieved from https://ec.europa.eu/eurostat/cros/content/composite-estimators-small-area-estimation-method_en.

EURAREA (2004). *Technical report.* EURAREA consortium. Project reference volume, deliverable d7.1.4. Retrieved from http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/spatial-analysis-andmodelling/eurarea/index.html.

European Commission Social Protection Committee (2001). *Report on indicators in the field of poverty and social exclusion.* Technical report.

European Commission (2010). *EUROPE 2020: A Strategy for smart, sustainable and inclusive growth.* Brussels.

European Council (2010). *European Council 17 June 2010 conclusions*, Brussels: European Council.

Eurostat (2011). *2009 Comparative EU intermediate quality report.* Version 3, Luxembourg: European Commission.

Eurostat (2012). *Quality report of the European Union Labour Force Survey 2010.* Methodologies & Working papers. European Union, 2012.

Eurostat (2013). *Methodological guidelines and description of EU-SILC target variables.* 2014 operation (Version September 2013).

Eurostat (2015). *EU statistics on income and living conditions (EU-SILC), Access to microdata.* Retrieved from http://ec.europa.eu/eurostat/web/microdata/european-unionstatistics-on-income-and-living-conditions.

Eurostat (2016). *EU statistics on income and living conditions (EU-SILC), methodology–data collection.* Retrieved from http://ec.europa.eu/ eurostat/statisticsexplained/ index.php/ EU_statistics_on_income_and_living_ conditions_(EU-SILC)_methodology_– _data collection.

Eurostat (2018). *Regions in the European Union. Nomenclature of territorial units for statistics* - NUTS 2016/EU-28. European Union, 2018.

Eurostat (2019). *Guidelines on small area estimation for city statistics and other functional geographics. Manuals and guidelines.* European Union, 2019.

Eurostat (2020). *Quality report of the European Union Labour Force Survey 2018. Population and social conditions, Statistical reports.* European Union, 2020.

Falorsi, S., & Solari, F. (2014). Theme: Small Area Estimation. *Memobust handbook on methodology of modern business statistic.* Eurostat, 2014.

Fasulo, A. (2012). Chapter 2. *ESSnet on small area estimation*, 45-66.

Fay, R.E., & Herriot, R.A. (1979). Estimates of income for small places: An application of James–Stein procedures to Census data. *Journal of the American Statistical Association, 74*, 269–277.

Foster, J. (1998). Absolute versus relative poverty. *The American Economic Review, 88*(2), 335-341.

Foster, J., Greer, J. & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica, 52,* 761–766.

Fouarge, D., & Layte, R. (2005). Welfare regimes and poverty dynamics: The duration and recurrence of poverty spells in Europe. *Journal of Social Policy,* 34(3), 407–426.

Ghosh, M., & Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science, 9,* 55-93

Ghosh, M., Natarajan, K., Walter, L. A., & Kim, D.H. (1999). Hierarchical Bayes GLMs for the analysis of spatial data: An application to disease mapping. *Journal of Statistical Planning and Inference, 75*, 305–318.

Gibson H., Palivos T., & Tavlas G. (2014). The crisis in the euro area: An analytic overview. *Journal of Macroeconomics*, *39*B, 233-239.

Gonzalez, M.E. (1973). Use and evaluation of synthetic estimates. *Proceedings of the Social Statistics Section*. American Statistical Association, 33-36.

Gonzalez-Manteiga, W., Lombardia, M. J., Molina, I., Morales, D., & Santamaria, L. (2010). Small area estimation under Fay-Herriot models with nonparametric estimation of heteroscedasticity. *Statistical Modelling*, *10*(2), 215–239.

Grant, L., (2012). The Sources of unemployment. NPG Forum Paper. Retrieved from https://npg.org/wp-content/uploads/2013/07/TheSourcesOfUnemployment.pdf

Guadarrama, M., Molina, I., & Rao, J.N.K. (2014). A comparison of small area estimation methods for poverty mapping. *Statistics in Transition new series and Survey Methodology, 17*(1), 41-66.

Hagenaars, A.J.M. (1986). *The perception of poverty.* North-Holland, Amsterdam.

Hagenaars, A.J.M. & De Vos, K. (1988). The definition and measurement of poverty, *Journal of Human Resources, 23*(2), 211-221.

Hagenaars, A., De Vos, K., & Zaidi, M.A. (1994). *Poverty statistics in the late 1980s: Research based on micro-data.* European Commission. Luxembourg.

Han, B. (2013). Conditional Akaike information criterion in the Fay-Herriot Model. *Survey Methodology*, *11,* 53–67.

Harding, A., & Rahman, A. (2017). *Small area estimation and microsimulation modelling.* Boca Raton, FL: CRC Press, Taylor & Francis Group.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association, 72,* 320-340.

Harville, D.A. (1991). Comment. *Statistical Science, 6,* 35-39.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2003). *The elements of statistical learning*. Springer-Verlag, New York.

Heady, P., and Hennell, S. (2000). *Small area estimation and the ecological effect - applying standard theory in practical situations,* ONS Methodology.

Hellenic Statistical Authority (2001). *Concepts and definitions: Census of population-housing and buildings (2001).* Retrieved from: https://www.statistics.gr/ el/statistics/-/publication/SAM04/2001.

Hellenic Statistical Authority (2009). *Metadata in Euro-SDMX format (ESMS) (2009).* Retrieved from: https://www.statistics.gr/en/ statistics/-/publication/SFA10/ 2009.

Hellenic Statistical Authority (2009a). *Risk of poverty,* Press Release. Retrieved from: https://www.statistics.gr/en/statistics/-/publication/SFA10/2009.

Hellenic Statistical Authority (2009b). *Labour force survey quality report.* Retrieved from: https://www.statistics.gr/en/statistics/-/publication/SJO01/2009-Q4.

Hellenic Statistical Authority (2011). *2011 Population and housing Census: Single integrated metadata structure (SIMS).* Retrieved from: https://www.statistics. gr/en/2011-Census-pop-hous.

Hellenic Statistical Authority (2012). Statistics on income and living conditions (EU-SILC), 2011. *Intermediate quality report.* General directorate of statistical surveys, division of population and labour market statistics.

Hellenic Statistical Authority (2013). *Metadata in Euro-SDMX format* (ESMS). Retrieved from: https://www.statistics.gr/el/statistics/-/publication/ SFA10/ 2013.

Hellenic Statistical Authority (2013a). Press release: *Risk of poverty*. Retrieved from: https://www.statistics.gr/el/statistics/-/publication/SFA10/2013.

Hellenic Statistical Authority (2013b). *Labour force survey quality report.* Retrieved from: https://www.statistics.gr/en/statistics/-/publication/SJO01/2013-Q4.

Hellenic Statistical Authority (2014). Single integrated metadata structure (SIMS). *Population-housing Census 2011*. Retrieved from: https://www.statistics.gr/en/2011-Census-pop-hous.

Hellenic Statistical Authority (2015). *Final quality report* (HBS) (2015). Retrieved from: https://www.statistics.gr/en/statistics/-/publication/SFA05/2015.

Henderson, C.R. (1950). Estimation genetics parameters (Abstract). *The Annals of Mathematical Statistics, 21*, 309-310

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under selection model. *Biometrics, 31*, 423-447.

Hidiroglou, M., & You, Y. (2016). Comparison of unit level and area level small area estimators. *Survey Methodology*, 42(1), 41-61.

Hodges, J.S., & Sargent, D.J. (2001). Counting degrees of freedom in hierarchical and other richly parameterized models. *Biometrika*, *88*, 367-379.

Holmoy, A.M.K., & Thomsen, I. (1998). Combining data from surveys and administrative record systems. The Norwegian experience. *International Statistical Review*, *66*, 201–221.

Hussmanns, R. (2007). *Measurement of employment, unemployment and underemployment – Current international standards and issues in their application.* ILO Bureau of Statistics, 2007.

International Conference of Labour Statisticians (Thirteenth ICLS), (1982). *Resolution concerning statistics of the economically active population, employment, unemployment and underemployment;* in: Current international recommendations on labour statistics, 2000 edition, ILO, Geneva, 2000, pp 86-87.

International Conference of Labour Statisticians (Fourteenth ICLS), (1987). *Guidelines on the implications of employment promotion schemes on the measurement of employment and unemployment;* in: Current international recommendations on labour statistics, 2000 edition, ILO, Geneva, 2000, pp. 24-28.

International Conference of Labour Statisticians (Fifteenth ICLS), (1993). *Resolution concerning the international classification of status in employment (ICSE)*; in: Current international recommendations on labour statistics, 2000 edition, ILO, Geneva, 2000, pp. 20-23.

International Conference of Labour Statisticians (Sixteenth ICLS), (1998). *Resolution concerning the measurement of underemployment and inadequate employment* *situations;* in: Current international recommendations on labour statistics, 2000 edition, ILO, Geneva, 2000, pp. 29-31.

International Labour Organization (2013). Report II. *Statistics of work, employment and labour underutilization:* Report for discussion at the 19th International Conference of Labour Statisticians, Geneva, 2013.

International Labour Organization (2019). *Quick guide on interpreting the unemployment rate.* International Labour Office, ILO, Geneva, 2019.

Jiang, J., & Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test, 15,* 1-96.

Jiang, J., Rao J.S., Gu Z., & Nguyen, T. (2008). Fence methods for mixed model selection. *The Annals of Statistics, 36*(4), 1669-1692.

Kackar, R.N., & Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedure for mixed linear models. *Communications in Statistics, Series A, 10,* 1249-1261.

Kass, R.E., & Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association, 84,* 717-726.

Kordos, J., (2016). Development of small area estimation in official statistics. *Statistics in Transition new series and Survey Methodology, 17 (1),* 105-132.

Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B, 60*(1), 65–81.

Laderchi, C., Saith, R., & Stewart, F., (2003). Does it matter that we do not agree on the definition of poverty: A comparison of four approaches. *Oxford Development Studies, 31(3),* pp. 233-274.

Laird, N.M., & Louis, T.A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association, 82,* 739-750.

Lehtonen, R., & Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. In *Sample Surveys: Inference and Analysis* (D. Pfeffermann and C. R. Rao, Eds.). *Handbook of Statistics, 29B,* 219–249. North-Holland, Amsterdam.

López-Vizcaíno, E., Lombardía, M.J., & Morales, D., (2015). Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. *Journal of the Royal Statistical Society, Series A (Statistics in Society), 178*(3), 535-565.

MacGibbon, B., & Tomberlin, T.J. (1989). Small area estimation of proportions via empirical Bayes techniques. *Survey Methodology, 15,* 237-252.

Malec, D., Sedransk, J., Moriarity, C.L., & LeClere, F.B. (1997). Small area inference for binary variables in national health interview survey. *Journal of the American Statistical Association, 92,* 815-826.

Meindl, B., (2008). Estimating unemployment-rates for small areas – A simulation-Based Approach. *Austrian Journal of Statistics, 37*(3), 349–360.

Mitrakos, T., (2014). *Inequality, poverty and social welfare in Greece: distributional effects of austerity.* Working Papers 174, Bank of Greece.

Molina, I., & Morales, D. (2009). Small area estimation of poverty indicators. *Boletin de Estadistica e Investigacion Operativa, 25,* 218–225.

Molina, I., Salvati, N., & Pratesi, M., (2009). Bootstrap for estimating the MSE of the spatial EBLUP. *Computational Statistics*, *24,* 441–458.

Molina, I., & Rao, J., (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistic, 38,* 369-385.

Molina, I., & Marhuenda, Y., (2015). sae: An R package for small area estimation. *R Journal*, *7,* 81-98. doi:10.32614/RJ-2015-007.

Morris, C.N., (1983). Parametric empirical Bayes inference theory and applications. *Journal of the American Statistical Association*, *78,* 47-65.

Moura, F.A.S., & Migon, H.S., (2002). Bayesian spatial models for small area proportions. *Statistical Modelling, 2*(3), 183–201.

Müller, S., Scealy, J.L., & Welsh, A.H. (2013). Model selection in linear mixed models, *Statistical Science, 28,* 135–167.

Narayan, D. (2000). *Voices of the poor: Can anyone hear us?* Washington, D.C.: World Bank Publication.

Omrani, H., Gerber, P., & Bousch, P. (2009). *Model-based small area estimation with application to unemployment estimates.* World Academy of Science, Engineering and Technology International Journal of Civil and Environmental Engineering, 3(1), 2009.

Office for National Statistics, (2004). *Labour force survey user guide*, volume 6. Office for National Statistics (ONS). Retrieved from: http://www.statistics.gov.uk/ downloads/theme labour/Vol6.pdf .

Organisation for Economic Co-operation and Development, (2013). *The OEDC approach to measure and monitor income poverty across countries*. Working paper 17, Session 2: Data comparability, Geneva, Switzerland.

Organisation for Economic Co-operation and Development, (2013a). *OECD framework for statistics on the distribution of household income, consumption and wealth.* Paris: OECD Publishing.

Paul, S. (1991). On the measurement of unemployment. *Journal of Development Economics, 36,* 395-404.

Pfeffermann, D. (2002). Small area estimation: New developments and directions. *International Statistical Review, 70,* 125-143.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science, 28*(1), 1-134.

Pfeffermann, D., & Tiller, R., (2006). Small area estimation with state space models subject to benchmark constraints. *Journal of the American Statistical Association*, *101,* 1387–1897.

Pfeffermann, D., & Sverchkov, M. (2007). Small area estimation under informative probability sampling of areas and within the selected areas. *Journal of the American Statistical Association*, *102*(480), 1427-1439.

Pigou, A.C. (1912). *Wealth and welfare.* New York: Macmillan.

Prasad, N.G.N., & Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, *85*(409), 163-171.

Prasad, N.G.N., & Rao, J.N.K. (1999). On robust small area estimation using a simple random effects model, *Survey Methodology*, 25(1), 67-72.

Pratesi, M. (2016). *Analysis of poverty data by small area estimation*. JohnWiley&Sons.

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.

Purcell, N.J., & Kish, L. (1980). Postcensal estimates for local areas (or domains). *International Statistical Review, 48,* 3-18.

Rahman, A. (2008). *A review of small area estimation problems and methodological developments.* University of Canberra, Canberra, Australia.

Rao, C.R., & Wu, Y. (2001). On model selection. In *Institute of Mathematical Statistics Lecture Notes-Monograph Series, 38,* 1-57. Institute of Mathematical Statistics, Beachwood, OH.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology, 25,* 175-186.

Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, Hoboken, New Jersey.

Rao, J.N.K., & 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., & Molina, I. (2015). *Small area estimation*, 2nd Edition. Wiley, Hoboken, New Jersey.

Ravallion, M., & Chen, S. (2011). Weakly relative poverty. *The Review of Economics and Statistics,* 93(4), 1251-1261.

Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science, 6*, 15-51.

Saei A., & Chambers R. (2003). Small area estimation: A review of methods based on the application of mixed models. *S3RI Methodology Working Papers*, M03/16. Southampton Statistical Sciences Research Institute, Southampton, UK.

Salvati, N., Giusti, C., & Pratesi, M., (2014). *The use of spatial information for the estimation of poverty indicators at the small area level.* In Betti, G., Lemmi, H., (Eds.) Poverty and Social Exclusion, New Methods of Analysis, Routledge, London, United Kingdom.

SAMPLE project, (2009). *New indicators and models for inequality and poverty with attention to social exclusion, vulnerability and deprivation.* Literature review, Volume I.

Sarndal, C. (1984). Design-consistent versus model-dependent estimation for small domains. *Journal of the American Statistical Association*, *79*(387), 624-631.

Sardnal, C.E., Swensson, B., & Wretman, J.H. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.

Schaible, W.L. (1996). *Indirect estimation in U.S. federal programs*. New York: Springer.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.

Sen, A.K. (1975). *Employment, technology and development.* Clarendon Press, Oxford.

Sen, A.K. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, *44*(2), 219-231.

Sen, A.K. (1983). Poor, relatively speaking. *Oxford Economic Papers, 35*(2), 153-169.

Sen, A.K. (1984). Rights and capabilities. In A.K. Sen, *Resources, Values and Development.* MA: Harvard University Press, pp 307-324.

Sen, A.K. (1985). *Commodities and capabilities.* North-Holland, Amsterdam.

Sen, A.K. (1992). *Inequality reexamined.* Cambridge: Harvard University Press.

Sengenberger, W. (2011). *Beyond the measurement of unemployment and underemployment: The case for extending and amending labour market statistics.* International Labour Office (ILO), Geneva.

Singh, M.P., Gambino, J.G., & Mantel, H. (1993). *Issues and options in the provision of small area statistics.* In G. Kalton, J. Kordos, and R. Platek (Eds.), Proceedings of the International Scientific Conference on Small Area Statistics and Survey Designs, 1, Central Statistical Office, Warsaw, 37–75.

Singh, M.P., Gambino, J., & Mantel, H.J. (1994). Issues and strategies for small area data. *Survey Methodology*, *20,* 3-14.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the measurement of economic performance and social progress* (pp. 292). Paris.

Stukel, D.M., & Rao, J.N.K. (1999). Small area estimation under two-fold nested errors regression models. *Journal of Statistical Planning and Inference*, *78,* 131-147.

Szymkowiak, M., Młodak, A., & Wawrowski, L. (2017). Mapping poverty at the level of subregions in Poland using indirect estimation. *Statistics in Transition new series, 18*(4), 609–635.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society B, 58,* 267–288.

Tiller, R.B. (1991). Times series modelling of sample survey data from the U.S. Current population survey. *Journal of Official Statistics*, *8,* 149-166.

Townsend, P. (1979). *Poverty in the United Kingdom. A survey of household resources and standards of living.* Middlesex: Penguin Books, 1216p.

Tzavidis, N., Salvati, N., Pratesi, M., & Chamber, R. (2008). M-Quantile models with application to poverty mapping. *Statistical Methods and Applications*, *17*(3), 393-411.

Tzavidis, N., Marchetti, S. & Chambers, R. (2010). Robust estimation of small area means and quantiles. *Australian and New Zealand Journal of Statistics*, *52,* 167–186.

Tzavidis, N., Zhang, L.-C., Luna, A., Schmid, T. & Rojas-Perilla, N. (2018). From start to finish: A framework for the production of small area official statistics. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), *181* (4), 927–979.

Ugarte1, M.D., Goicoa, T., Militino, A.F., & Sagaseta-Lopez, M. (2009). Estimating unemployment in very small areas. *Statistics and Operations Research Transactions, 33* (1), 49-70.

UNESCO, (2006). *International standard classification of education (ISCED) 1997*, Re-edition, May 2006.

United Nations Economic Commission for Europe-UNECE, (2011). *Canberra group handbook on household income statistics:* Second Edition. Geneva: United Nations.

United Nations Economic Commission for Europe-UNECE, (2017). *Guide on poverty measurement*, New York and Geneva: United Nations. ECE/CES/STAT/2017 /4.

Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, *92,* 351-370.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, *54* (3), 426–482.

Wang, J., Fuller, W.A., & Qu, Y. (2008). Small area estimation under a restriction. *Survey Methodology*, *34,* 1, 29–36.

Whitworth, A. (2013). *Evaluations and improvements in small area estimation methodologies.* Discussion Paper. NCRM.

Wolter, K.M. (2007). *Introduction to variance estimation*. Springer, London.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika*, *92*(4), 937-950.

You, Y., & 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. (2002a). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Canadian Journal of Statistics*, *30,* 431-439.

You, Y., & Rao, J.N.K. (2002b). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, *30,* 3-15.

You, Y., & Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology*, *32*, 97–103.

Appendix

Table A1 Correlation matrix of the initial set of auxiliary variables from the 2011 Census in Greece

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.00	-0.91	-0.83	0.98	-0.79	-0.26	-0.73	0.60	0.63	-0.26
X2	-0.91	1.00	0.54	-0.93	0.49	0.16	0.68	-0.50	-0.52	0.06
X3	-0.83	0.54	1.00	-0.78	0.98	0.10	0.48	-0.39	-0.43	0.45
X4	0.98	-0.93	-0.78	1.00	-0.74	-0.22	-0.73	0.57	0.59	-0.15
X5	-0.79	0.49	0.98	-0.74	1.00	0.14	0.48	-0.38	-0.42	0.54
X6	-0.26	0.16	0.10	-0.22	0.14	1.00	0.63	-0.84	-0.80	0.13
X7	-0.73	0.68	0.48	-0.73	0.48	0.63	1.00	-0.84	-0.87	0.17
X8	0.60	-0.50	-0.39	0.57	-0.38	-0.84	-0.84	1.00	0.98	-0.16
X9	0.63	-0.52	-0.43	0.59	-0.42	-0.80	-0.87	0.98	1.00	-0.21
X10	-0.26	0.06	0.45	-0.15	0.54	0.13	0.17	-0.16	-0.21	1.00
X11	-0.71	0.54	0.72	-0.67	0.74	0.23	0.55	-0.44	-0.47	0.57
X12	0.29	-0.32	-0.19	0.36	-0.11	0.07	-0.16	0.12	0.10	0.26
X13	0.70	-0.66	-0.47	0.73	-0.47	-0.49	-0.90	0.74	0.79	-0.16
X14	0.08	-0.25	0.19	0.11	0.27	-0.06	-0.11	0.11	0.10	0.33
X15	-0.66	0.35	0.94	-0.59	0.93	0.00	0.25	-0.22	-0.25	0.48
X16	0.85	-0.78	-0.74	0.81	-0.72	-0.02	-0.42	0.23	0.24	-0.23
X17	-0.31	0.29	0.19	-0.29	0.16	0.29	0.27	-0.39	-0.35	-0.09
X18	0.37	-0.28	-0.38	0.33	-0.38	-0.16	-0.42	0.26	0.34	-0.21
X19	0.47	-0.48	-0.34	0.47	-0.36	-0.08	-0.34	0.10	0.15	-0.01
X20	-0.25	0.38	-0.04	-0.31	-0.07	0.21	0.35	-0.32	-0.33	-0.17
X21	0.36	-0.32	-0.32	0.36	-0.37	-0.12	-0.27	0.14	0.18	-0.19
X22	0.82	-0.64	-0.75	0.77	-0.75	-0.51	-0.81	0.75	0.79	-0.42
X23	0.77	-0.69	-0.64	0.72	-0.60	-0.24	-0.56	0.49	0.53	-0.24
X24	-0.07	0.20	-0.09	-0.14	-0.14	-0.12	-0.06	0.11	0.09	-0.41
X25	-0.72	0.64	0.58	-0.73	0.56	0.33	0.62	-0.66	-0.67	0.23
X26	0.29	-0.18	-0.38	0.25	-0.42	0.00	-0.17	0.00	0.01	-0.30
X27	0.08	-0.01	-0.25	0.04	-0.18	0.46	0.01	-0.28	-0.22	-0.19
X28	0.39	-0.34	-0.44	0.37	-0.37	0.30	-0.16	-0.01	0.01	-0.21
X29	0.06	-0.16	0.05	0.13	0.09	0.17	-0.04	-0.06	-0.04	0.21
X30	0.70	-0.65	-0.46	0.73	-0.45	-0.60	-0.87	0.84	0.86	-0.11
X31	-0.44	0.55	0.26	-0.47	0.23	-0.32	0.14	0.04	0.00	0.09
X32	0.50	-0.51	-0.24	0.55	-0.21	-0.47	-0.66	0.50	0.54	0.26

Table A1 Continued

	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
X1	-0.71	0.29	0.70	0.08	-0.66	0.85	-0.31	0.37	0.47	-0.25
X2	0.54	-0.32	-0.66	-0.25	0.35	-0.78	0.29	-0.28	-0.48	0.38
X3	0.72	-0.19	-0.47	0.19	0.94	-0.74	0.19	-0.38	-0.34	-0.04
X4	-0.67	0.36	0.73	0.11	-0.59	0.81	-0.29	0.33	0.47	-0.31
X5	0.74	-0.11	-0.47	0.27	0.93	-0.72	0.16	-0.38	-0.36	-0.07
X6	0.23	0.07	-0.49	-0.06	0.00	-0.02	0.29	-0.16	-0.08	0.21
X7	0.55	-0.16	-0.90	-0.11	0.25	-0.42	0.27	-0.42	-0.34	0.35
X8	-0.44	0.12	0.74	0.11	-0.22	0.23	-0.39	0.26	0.10	-0.32
X9	-0.47	0.10	0.79	0.10	-0.25	0.24	-0.35	0.34	0.15	-0.33
X10	0.57	0.26	-0.16	0.33	0.48	-0.23	-0.09	-0.21	-0.01	-0.17
X11	1.00	-0.28	-0.66	0.05	0.58	-0.57	0.02	-0.28	-0.22	0.29
X12	-0.28	1.00	0.36	0.42	-0.05	0.14	0.00	-0.26	-0.15	-0.50
X13	-0.66	0.36	1.00	0.22	-0.19	0.34	-0.19	0.30	0.26	-0.53
X14	0.05	0.42	0.22	1.00	0.32	0.00	-0.10	-0.22	-0.12	-0.51
X15	0.58	-0.05	-0.19	0.32	1.00	-0.71	0.15	-0.38	-0.33	-0.23
X16	-0.57	0.14	0.34	0.00	-0.71	1.00	-0.23	0.42	0.67	-0.04
X17	0.02	0.00	-0.19	-0.10	0.15	-0.23	1.00	-0.07	-0.09	0.10
X18	-0.28	-0.26	0.30	-0.22	-0.38	0.42	-0.07	1.00	0.72	0.21
X19	-0.22	-0.15	0.26	-0.12	-0.33	0.67	-0.09	0.72	1.00	0.04
X20	0.29	-0.50	-0.53	-0.51	-0.23	-0.04	0.10	0.21	0.04	1.00
X21	-0.28	-0.07	0.21	-0.27	-0.33	0.47	0.00	0.54	0.63	0.15
X22	-0.75	0.11	0.76	-0.01	-0.62	0.60	-0.26	0.53	0.37	-0.17
X23	-0.58	0.02	0.50	0.06	-0.56	0.76	-0.30	0.48	0.52	-0.23
X24	-0.18	-0.33	0.04	-0.32	-0.11	-0.12	0.12	0.25	-0.04	0.34
X25	0.74	-0.55	-0.75	-0.15	0.39	-0.42	0.17	-0.09	-0.08	0.51
X26	-0.37	-0.10	0.09	-0.37	-0.47	0.48	0.09	0.49	0.51	0.34
X27	-0.38	0.22	0.10	-0.06	-0.22	0.14	0.30	0.27	0.03	0.06
X28	-0.60	0.44	0.30	0.09	-0.35	0.38	0.15	0.10	0.03	-0.21
X29	0.14	0.08	0.09	0.22	0.13	-0.03	0.02	-0.31	-0.10	-0.20
X30	-0.56	0.48	0.90	0.22	-0.23	0.33	-0.26	0.23	0.19	-0.48
X31	0.08	0.03	-0.18	-0.17	0.15	-0.39	0.15	0.01	-0.21	0.15
X32	-0.26	0.30	0.65	0.25	-0.04	0.30	-0.21	0.25	0.36	-0.34

Table A1 Continued

	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30	X31	X32
X1	0.36	0.82	0.77	-0.07	-0.72	0.29	0.08	0.39	0.06	0.70	-0.44	0.50
X2	-0.32	-0.64	-0.69	0.20	0.64	-0.18	-0.01	-0.34	-0.16	-0.65	0.55	-0.51
X3	-0.32	-0.75	-0.64	-0.09	0.58	-0.38	-0.25	-0.44	0.05	-0.46	0.26	-0.24
X4	0.36	0.77	0.72	-0.14	-0.73	0.25	0.04	0.37	0.13	0.73	-0.47	0.55
X5	-0.37	-0.75	-0.60	-0.14	0.56	-0.42	-0.18	-0.37	0.09	-0.45	0.23	-0.21
X6	-0.12	-0.51	-0.24	-0.12	0.33	0.00	0.46	0.30	0.17	-0.60	-0.32	-0.47
X7	-0.27	-0.81	-0.56	-0.06	0.62	-0.17	0.01	-0.16	-0.04	-0.87	0.14	-0.66
X8	0.14	0.75	0.49	0.11	-0.66	0.00	-0.28	-0.01	-0.06	0.84	0.04	0.50
X9	0.18	0.79	0.53	0.09	-0.67	0.01	-0.22	0.01	-0.04	0.86	0.00	0.54
X10	-0.19	-0.42	-0.24	-0.41	0.23	-0.30	-0.19	-0.21	0.21	-0.11	0.09	0.26
X11	-0.28	-0.75	-0.58	-0.18	0.74	-0.37	-0.38	-0.60	0.14	-0.56	0.08	-0.26
X12	-0.07	0.11	0.02	-0.33	-0.55	-0.10	0.22	0.44	0.08	0.48	0.03	0.30
X13	0.21	0.76	0.50	0.04	-0.75	0.09	0.10	0.30	0.09	0.90	-0.18	0.65
X14	-0.27	-0.01	0.06	-0.32	-0.15	-0.37	-0.06	0.09	0.22	0.22	-0.17	0.25
X15	-0.33	-0.62	-0.56	-0.11	0.39	-0.47	-0.22	-0.35	0.13	-0.23	0.15	-0.04
X16	0.47	0.60	0.76	-0.12	-0.42	0.48	0.14	0.38	-0.03	0.33	-0.39	0.30
X17	0.00	-0.26	-0.30	0.12	0.17	0.09	0.30	0.15	0.02	-0.26	0.15	-0.21
X18	0.54	0.53	0.48	0.25	-0.09	0.49	0.27	0.10	-0.31	0.23	0.01	0.25
X19	0.63	0.37	0.52	-0.04	-0.08	0.51	0.03	0.03	-0.10	0.19	-0.21	0.36
X20	0.15	-0.17	-0.23	0.34	0.51	0.34	0.06	-0.21	-0.20	-0.48	0.15	-0.34
X21	1.00	0.34	0.32	0.04	-0.19	0.49	0.02	0.12	-0.06	0.24	-0.13	0.15
X22	0.34	1.00	0.67	0.20	-0.68	0.40	0.15	0.36	-0.19	0.74	-0.06	0.43
X23	0.32	0.67	1.00	-0.26	-0.55	0.14	0.11	0.37	-0.01	0.43	-0.30	0.34
X24	0.04	0.20	-0.26	1.00	0.08	0.49	0.19	-0.01	-0.29	-0.01	0.20	-0.22
X25	-0.19	-0.68	-0.55	0.08	1.00	-0.06	-0.12	-0.55	0.01	-0.78	0.13	-0.27
X26	0.49	0.40	0.14	0.49	-0.06	1.00	0.22	0.15	-0.26	0.11	0.15	0.06
X27	0.02	0.15	0.11	0.19	-0.12	0.22	1.00	0.80	-0.22	-0.12	0.11	-0.10
X28	0.12	0.36	0.37	-0.01	-0.55	0.15	0.80	1.00	-0.13	0.14	-0.01	-0.09
X29	-0.06	-0.19	-0.01	-0.29	0.01	-0.26	-0.22	-0.13	1.00	0.09	-0.43	0.22
X30	0.24	0.74	0.43	-0.01	-0.78	0.11	-0.12	0.14	0.09	1.00	-0.11	0.67
X31	-0.13	-0.06	-0.30	0.20	0.13	0.15	0.11	-0.01	-0.43	-0.11	1.00	-0.10
X32	0.15	0.43	0.34	-0.22	-0.27	0.06	-0.10	-0.09	0.22	0.67	-0.10	1.00

Table A2 Correlation matrix of the initial set of auxiliary variables from the 2001 Census in Greece

	X1	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**	**X8**	**X9**	**X10**
X1	1.00	-0.96	-0.84	0.97	-0.77	0.01	-0.58	0.51	0.48	-0.15
X2	-0.96	1.00	0.66	-0.96	0.59	0.03	0.59	-0.53	-0.48	0.02
X3	-0.84	0.66	1.00	-0.78	0.97	-0.09	0.42	-0.36	-0.38	0.38
X4	0.97	-0.96	-0.78	1.00	-0.72	-0.08	-0.62	0.57	0.51	-0.01
X5	-0.77	0.59	0.97	-0.72	1.00	-0.02	0.47	-0.38	-0.41	0.49
X6	0.01	0.03	-0.09	-0.08	-0.02	1.00	0.41	-0.74	-0.68	-0.08
X7	-0.58	0.59	0.42	-0.62	0.47	0.41	1.00	-0.74	-0.75	0.21
X8	0.51	-0.53	-0.36	0.57	-0.38	-0.74	-0.74	1.00	0.95	0.00
X9	0.48	-0.48	-0.38	0.51	-0.41	-0.68	-0.75	0.95	1.00	-0.13
X10	-0.15	0.02	0.38	-0.01	0.49	-0.08	0.21	0.00	-0.13	1.00
X11	-0.21	0.09	0.41	-0.09	0.52	0.00	0.33	-0.10	-0.23	0.98
X12	0.09	-0.05	-0.15	0.06	-0.10	0.26	0.24	-0.22	-0.20	0.02
X13	0.24	-0.23	-0.20	0.33	-0.30	-0.46	-0.66	0.61	0.62	-0.15
X14	0.31	-0.32	-0.23	0.29	-0.34	-0.10	-0.47	0.27	0.41	-0.37
X15	0.13	-0.08	-0.18	0.08	-0.28	-0.02	-0.33	0.16	0.29	-0.43
X16	0.35	-0.37	-0.24	0.34	-0.31	-0.10	-0.57	0.32	0.45	-0.29
X17	0.21	-0.20	-0.18	0.19	-0.25	-0.08	-0.37	0.20	0.32	-0.30
X18	-0.02	0.11	-0.17	-0.09	-0.22	0.38	0.18	-0.41	-0.34	-0.29
X19	-0.79	0.75	0.69	-0.80	0.64	0.16	0.61	-0.64	-0.57	0.14

Table A2 Continued

	X11	**X12**	**X13**	**X14**	**X15**	**X16**	**X17**	**X18**	**X19**
X1	-0.21	0.09	0.24	0.31	0.13	0.35	0.21	-0.02	-0.79
X2	0.09	-0.05	-0.23	-0.32	-0.08	-0.37	-0.20	0.11	0.75
X3	0.41	-0.15	-0.20	-0.23	-0.18	-0.24	-0.18	-0.17	0.69
X4	-0.09	0.06	0.33	0.29	0.08	0.34	0.19	-0.09	-0.80
X5	0.52	-0.10	-0.30	-0.34	-0.28	-0.31	-0.25	-0.22	0.64
X6	0.00	0.26	-0.46	-0.10	-0.02	-0.10	-0.08	0.38	0.16
X7	0.33	0.24	-0.66	-0.47	-0.33	-0.57	-0.37	0.18	0.61
X8	-0.10	-0.22	0.61	0.27	0.16	0.32	0.20	-0.41	-0.64
X9	-0.23	-0.20	0.62	0.41	0.29	0.45	0.32	-0.34	-0.57
X10	0.98	0.02	-0.15	-0.37	-0.43	-0.29	-0.30	-0.29	0.14
X11	1.00	0.06	-0.23	-0.42	-0.47	-0.36	-0.35	-0.24	0.19
X12	0.06	1.00	0.05	-0.25	-0.24	-0.25	-0.09	-0.07	-0.03
X13	-0.23	0.05	1.00	0.15	0.08	0.19	0.22	-0.33	-0.46
X14	-0.42	-0.25	0.15	1.00	0.81	0.92	0.73	0.20	-0.07
X15	-0.47	-0.24	0.08	0.81	1.00	0.81	0.69	0.18	-0.07
X16	-0.36	-0.25	0.19	0.92	0.81	1.00	0.76	0.03	-0.12
X17	-0.35	-0.09	0.22	0.73	0.69	0.76	1.00	-0.15	0.00
X18	-0.24	-0.07	-0.33	0.20	0.18	0.03	-0.15	1.00	0.08
X19	0.19	-0.03	-0.46	-0.07	-0.07	-0.12	0.00	0.08	1.00

Table A3 Direct estimates of headcount ratio in Greece for the year 2013 and the corresponding sample sizes, variances (SD^2), standard errors (SD) and coefficients of variance (CV)

Nomos	Nomos_name	Sample_Size	Direct_Headcount-Ratio	SD	SD^2	CV
300001	Etolia and Akarnania	410	0.401382	0.036449	0.001329	9.080926
300003	Viotia	143	0.125327	0.028251	0.000798	22.5417
300004	Evia	341	0.277133	0.032941	0.001085	11.88621
300005	Evrytania	71	0.195504	0.052933	0.002802	27.07525
300006	Fthiotida	242	0.197842	0.030702	0.000943	15.51837
300007	Fokida	128	0.051831	0.01875	0.000352	36.17431
300011	Argolida	146	0.307691	0.048982	0.002399	15.91937
300012	Arkadia	206	0.202858	0.032785	0.001075	16.16178
300013	Achaia	794	0.272157	0.019959	0.000398	7.333704
300014	Ilia	242	0.210363	0.032254	0.00104	15.33247
300015	Korinthia	373	0.251297	0.028902	0.000835	11.50095
300016	Lakonia	63	0.211938	0.055386	0.003068	26.133
300017	Messinia	333	0.261845	0.03423	0.001172	13.0725
300021	Zakynthos	106	0.280742	0.064222	0.004125	22.87598
300022	Kerkyra	113	0.196137	0.046391	0.002152	23.65214
300023	Kefallinia	69	0.380932	0.077898	0.006068	20.44937
300024	Lefkada	40	0.187066	0.062689	0.00393	33.51147
300031	Arta	126	0.260256	0.049419	0.002442	18.98854
300032	Thesprotia	31	0.571288	0.17239	0.029718	30.1756
300033	Loannina	417	0.246968	0.028888	0.000835	11.69714
300034	Preveza	178	0.200535	0.037747	0.001425	18.82332
300041	Karditsa	269	0.215052	0.029087	0.000846	13.52577
300042	Larissa	563	0.197767	0.019615	0.000385	9.918352
300043	Magnissia	430	0.227893	0.027369	0.000749	12.00981
300044	Trikala	83	0.180892	0.04742	0.002249	26.2144
300051	Grevena	37	0.270485	0.095891	0.009195	35.45158
300052	Drama	104	0.269402	0.056295	0.003169	20.89617
300053	Imathia	292	0.463332	0.044141	0.001948	9.526753
300054	Thessaloniki	1520	0.221002	0.013733	0.000189	6.213892
300055	Kavala	255	0.18601	0.033391	0.001115	17.95107
300056	Kastoria	135	0.249433	0.050008	0.002501	20.04867
300057	Kilkis	165	0.199317	0.035475	0.001258	17.79833
300058	Kozani	268	0.242074	0.034979	0.001224	14.44957
300059	Pella	199	0.341688	0.046218	0.002136	13.52628
300061	Pieria	370	0.302909	0.03146	0.00099	10.38598
300062	Serres	327	0.234048	0.027409	0.000751	11.71104
300063	Florina	106	0.272819	0.05278	0.002786	19.34606
300064	Chalkidiki and Aghion Oros	120	0.285653	0.054174	0.002935	18.96499
300071	Evros	332	0.210024	0.026885	0.000723	12.80086
300072	Xanthi	173	0.380403	0.04838	0.002341	12.71815
300073	Rodopi	230	0.355588	0.052988	0.002808	14.90136
300081	Dodekanissos	253	0.269381	0.035233	0.001241	13.07939
300082	Kyklades	173	0.191467	0.034609	0.001198	18.07564
300083	Lesvos	295	0.289127	0.033027	0.001091	11.42304
300084	Samos	28	0.443847	0.181398	0.032905	40.86957
300085	Chios	62	0.23003	0.067912	0.004612	29.52323
300091	Iraklio	622	0.198484	0.019958	0.000398	10.05538
300092	Lassithi	56	0.340881	0.08693	0.007557	25.5016
300093	Rethymno	125	0.284306	0.06058	0.00367	21.30792
300094	Chania	244	0.197354	0.031679	0.001004	16.05168
300101	Prefecture of Athens	3873	0.201803	0.010519	0.000111	5.212367
300102	Prefecture of East Attiki	871	0.223703	0.018537	0.000344	8.286498
300103	Prefecture of West Attiki	226	0.340047	0.042185	0.00178	12.40557
300104	Prefecture of Pireas	652	0.141999	0.015676	0.000246	11.03985

Table A4 Direct estimates of poverty gap in Greece for the year 2013 and the corresponding sample sizes, variances (SD^2), standard errors (SD) and coefficients of variance (CV)

Nomos	Nomos_name	Sample_Size	Direct_Poverty-Gap	SD	SD^2	CV
300001	Etolia and Akarnania	410	0.142915	0.014751	0.000218	10.32139
300003	Viotia	143	0.017136	0.004387	1.92E-05	25.60004
300004	Evia	341	0.124678	0.020147	0.000406	16.15914
300005	Evrytania	71	0.045799	0.01778	0.000316	38.82274
300006	Fthiotida	242	0.084714	0.015969	0.000255	18.85103
300007	Fokida	128	0.007021	0.003089	9.54E-06	44.00429
300011	Argolida	146	0.110253	0.021175	0.000448	19.20599
300012	Arkadia	206	0.064441	0.013804	0.000191	21.4215
300013	Achaia	794	0.120094	0.010926	0.000119	9.098143
300014	Ilia	242	0.057449	0.009748	9.50E-05	16.96729
300015	Korinthia	373	0.103305	0.013945	0.000194	13.499
300016	Lakonia	63	0.02844	0.008943	8.00E-05	31.44705
300017	Messinia	333	0.101922	0.015312	0.000234	15.02321
300021	Zakynthos	106	0.106228	0.025914	0.000672	24.39443
300022	Kerkyra	113	0.114849	0.032774	0.001074	28.53703
300023	Kefallinia	69	0.155102	0.031274	0.000978	20.16345
300024	Lefkada	40	0.019929	0.009289	8.63E-05	46.61149
300031	Arta	126	0.152025	0.035594	0.001267	23.41304
300032	Thesprotia	31	0.189766	0.055207	0.003048	29.09196
300033	Loannina	417	0.072885	0.010349	0.000107	14.19947
300034	Preveza	178	0.100124	0.021761	0.000474	21.73402
300041	Karditsa	269	0.093209	0.016096	0.000259	17.26851
300042	Larissa	563	0.096378	0.011427	0.000131	11.85612
300043	Magnissia	430	0.073229	0.009406	8.85E-05	12.84404
300044	Trikala	83	0.058331	0.015995	0.000256	27.42069
300051	Grevena	37	0.052448	0.020872	0.000436	39.79517
300052	Drama	104	0.121536	0.027469	0.000755	22.6017
300053	Imathia	292	0.24384	0.027858	0.000776	11.42459
300054	Thessaloniki	1520	0.080789	0.006297	3.97E-05	7.794411
300055	Kavala	255	0.098087	0.02162	0.000467	22.04156
300056	Kastoria	135	0.076452	0.016178	0.000262	21.16152
300057	Kilkis	165	0.075511	0.014591	0.000213	19.32284
300058	Kozani	268	0.072067	0.011493	0.000132	15.94775
300059	Pella	199	0.113593	0.020357	0.000414	17.92147
300061	Pieria	370	0.120276	0.01376	0.000189	11.44045
300062	Serres	327	0.077677	0.01185	0.00014	15.2557
300063	Florina	106	0.103116	0.020958	0.000439	20.32482
300064	Chalkidiki and Aghion Oros	120	0.062297	0.014702	0.000216	23.59991
300071	Evros	332	0.064245	0.010447	0.000109	16.2614
300072	Xanthi	173	0.180456	0.026992	0.000729	14.95756
300073	Rodopi	230	0.096135	0.01433	0.000205	14.90574
300081	Dodekanissos	253	0.069444	0.011505	0.000132	16.56697
300082	Kyklades	173	0.058281	0.015537	0.000241	26.65856
300083	Lesvos	295	0.105476	0.013294	0.000177	12.60396
300084	Samos	28	0.23757	0.123642	0.015287	52.04453
300085	Chios	62	0.091433	0.035431	0.001255	38.75082
300091	Iraklio	622	0.071539	0.008597	7.39E-05	12.01654
300092	Lassithi	56	0.180696	0.05374	0.002888	29.74025
300093	Rethymno	125	0.121115	0.032311	0.001044	26.67824
300094	Chania	244	0.054706	0.010315	0.000106	18.85534
300101	Prefecture of Athens	3873	0.072704	0.003907	1.53E-05	5.374239
300102	Prefecture of East Attiki	871	0.074739	0.00773	5.97E-05	10.34217
300103	Prefecture of West Attiki	226	0.128626	0.018056	0.000326	14.03757
300104	Prefecture of Pireas	652	0.058523	0.008045	6.47E-05	13.74734

Table A5 Direct estimates of headcount ratio in Greece for the year 2009 and the corresponding sample sizes, variances (SD^2), standard errors (SD) and coefficients of variance (CV)

Nomos	Nomos_name	Sample_Size	Direct_Headcount-ratio	SD	SD^2	CV
300001	Etolia and Akarnania	278	0.238217	0.038245	0.001463	16.05461
300003	Viotia	184	0.158318	0.033468	0.00112	21.13958
300004	Evia	254	0.397155	0.054367	0.002956	13.68907
300005	Evrytania	35	0.456709	0.145237	0.021094	31.8007
300006	Fthiotida	355	0.205888	0.034212	0.00117	16.61672
300007	Fokida	82	0.417272	0.111141	0.012352	26.63507
300011	Argolida	201	0.27051	0.061572	0.003791	22.76152
300012	Arkadia	171	0.270788	0.0468	0.00219	17.28295
300013	Achaia	573	0.264978	0.03147	0.00099	11.87654
300014	Ilia	369	0.363437	0.039158	0.001533	10.77432
300015	Korinthia	333	0.164094	0.026592	0.000707	16.20506
300016	Lakonia	123	0.277055	0.050738	0.002574	18.31336
300017	Messinia	220	0.47012	0.097606	0.009527	20.76183
300021	Zakynthos	31	0.479406	0.14891	0.022174	31.06146
300022	Kerkyra	183	0.107219	0.026093	0.000681	24.33622
300024	Lefkada	28	0.255193	0.101933	0.01039	39.9437
300031	Arta	118	0.344103	0.075568	0.00571	21.96074
300032	Thesprotia	59	0.288542	0.103311	0.010673	35.80447
300033	Loannina	204	0.219443	0.046701	0.002181	21.2817
300034	Preveza	106	0.393241	0.072783	0.005297	18.50854
300041	Karditsa	387	0.178901	0.023158	0.000536	12.94469
300042	Larissa	488	0.197933	0.029361	0.000862	14.83369
300043	Magnissia	346	0.144732	0.023528	0.000554	16.25605
300044	Trikala	127	0.234552	0.053603	0.002873	22.85331
300051	Grevena	67	0.417528	0.100017	0.010003	23.95455
300052	Drama	215	0.316476	0.053452	0.002857	16.88983
300053	Imathia	305	0.251217	0.037978	0.001442	15.11748
300054	Thessaloniki	1746	0.182245	0.01303	0.00017	7.149477
300055	Kavala	418	0.2765	0.032676	0.001068	11.81771
300056	Kastoria	98	0.302042	0.069861	0.004881	23.12948
300057	Kilkis	282	0.470629	0.064461	0.004155	13.69669
300058	Kozani	226	0.1481	0.026852	0.000721	18.13071
300059	Pella	261	0.268012	0.032486	0.001055	12.12105
300061	Pieria	315	0.328186	0.040119	0.00161	12.22444
300062	Serres	343	0.481144	0.050912	0.002592	10.58152
300063	Florina	147	0.374037	0.06793	0.004614	18.16125
300064	Chalkidiki and Aghion Oros	101	0.23681	0.066697	0.004448	28.16474
300071	Evros	246	0.261304	0.044221	0.001955	16.92317
300072	Xanthi	102	0.465969	0.094752	0.008978	20.3343
300073	Rodopi	226	0.326095	0.059319	0.003519	18.19069
300081	Dodekanissos	330	0.228706	0.039042	0.001524	17.0708
300082	Kyklades	225	0.10435	0.026199	0.000686	25.10689
300083	Lesvos	238	0.450904	0.072492	0.005255	16.07694
300084	Samos	22	0.380571	0.190228	0.036187	49.98488
300085	Chios	148	0.185181	0.047455	0.002252	25.62617
300091	Iraklio	570	0.134347	0.021951	0.000482	16.33925
300092	Lassithi	140	0.090286	0.029484	0.000869	32.65671
300093	Rethymno	71	0.038365	0.020345	0.000414	53.0307
300094	Chania	339	0.193932	0.030666	0.00094	15.81254
300101	Prefecture of Athens	3934	0.133447	0.007672	5.89E-05	5.749034
300102	Prefecture of East Attiki	680	0.109822	0.015249	0.000233	13.88532
300103	Prefecture of West Attiki	298	0.22279	0.034264	0.001174	15.37971
300104	Prefecture of Pireas	670	0.154393	0.018466	0.000341	11.96058

Table A6 Direct estimates of poverty gap in Greece for the year 2009 and the corresponding sample sizes, variances (SD^2), standard errors (SD) and coefficients of variance (CV)

Nomos	Nomos_name	Sample_Size	Direct_Poverty-gap	SD	SD^2	CV
300001	Etolia and Akarnania	278	0.05752	0.010954	0.00012	19.04442
300003	Viotia	184	0.060157	0.015906	0.000253	26.44138
300004	Evia	254	0.177562	0.028654	0.000821	16.13757
300005	Evrytania	35	0.11566	0.034344	0.001179	29.69361
300006	Fthiotida	355	0.045058	0.008496	7.22E-05	18.85643
300007	Fokida	82	0.084115	0.018931	0.000358	22.50582
300011	Argolida	201	0.075954	0.018677	0.000349	24.58916
300012	Arkadia	171	0.114397	0.024466	0.000599	21.38712
300013	Achaia	573	0.072607	0.011283	0.000127	15.53982
300014	Ilia	369	0.118334	0.01506	0.000227	12.72694
300015	Korinthia	333	0.07062	0.014911	0.000222	21.11462
300016	Lakonia	123	0.081342	0.018246	0.000333	22.43069
300017	Messinia	220	0.155811	0.042493	0.001806	27.27206
300021	Zakynthos	31	0.095775	0.027528	0.000758	28.74213
300022	Kerkyra	183	0.033385	0.008275	6.85E-05	24.78713
300024	Lefkada	28	0.076687	0.032342	0.001046	42.17408
300031	Arta	118	0.091473	0.024286	0.00059	26.55007
300032	Thesprotia	59	0.053528	0.018557	0.000344	34.66852
300033	Loannina	204	0.033497	0.007857	6.17E-05	23.45506
300034	Preveza	106	0.116287	0.024774	0.000614	21.30423
300041	Karditsa	387	0.065764	0.012488	0.000156	18.9886
300042	Larissa	488	0.088395	0.018564	0.000345	21.0012
300043	Magnissia	346	0.040158	0.007795	6.08E-05	19.40987
300044	Trikala	127	0.107087	0.030429	0.000926	28.41496
300051	Grevena	67	0.113839	0.031626	0.001	27.78105
300052	Drama	215	0.117326	0.025693	0.00066	21.89859
300053	Imathia	305	0.074229	0.014785	0.000219	19.91817
300054	Thessaloniki	1746	0.050604	0.004541	2.06E-05	8.972831
300055	Kavala	418	0.066857	0.011299	0.000128	16.90032
300056	Kastoria	98	0.108343	0.029464	0.000868	27.19497
300057	Kilkis	282	0.169787	0.025387	0.000644	14.95206
300058	Kozani	226	0.051568	0.011556	0.000134	22.40906
300059	Pella	261	0.090622	0.015354	0.000236	16.94293
300061	Pieria	315	0.078418	0.010199	0.000104	13.00562
300062	Serres	343	0.157912	0.020566	0.000423	13.02374
300063	Florina	147	0.134295	0.036722	0.001349	27.34433
300064	Chalkidiki and Aghion Oros	101	0.04756	0.014838	0.00022	31.19854
300071	Evros	246	0.072517	0.012602	0.000159	17.37824
300072	Xanthi	102	0.141702	0.035189	0.001238	24.833
300073	Rodopi	226	0.120702	0.025666	0.000659	21.26397
300081	Dodekanissos	330	0.049064	0.008855	7.84E-05	18.04738
300082	Kyklades	225	0.026676	0.005861	3.44E-05	21.97196
300083	Lesvos	238	0.085953	0.016339	0.000267	19.00878
300084	Samos	22	0.065275	0.032627	0.001065	49.98488
300085	Chios	148	0.042087	0.011665	0.000136	27.71681
300091	Iraklio	570	0.040799	0.007825	6.12E-05	19.17839
300092	Lassithi	140	0.035543	0.017686	0.000313	49.76012
300093	Rethymno	71	0.005642	0.002908	8.45E-06	51.54146
300094	Chania	339	0.052601	0.011883	0.000141	22.59052
300101	Prefecture of Athens	3934	0.031686	0.002312	5.34E-06	7.295638
300102	Prefecture of East Attiki	680	0.042888	0.007541	5.69E-05	17.58225
300103	Prefecture of West Attiki	298	0.086327	0.015201	0.000231	17.60821
300104	Prefecture of Pireas	670	0.04746	0.007187	5.17E-05	15.1442

Table A7 EBLUP Fay and Herriot estimates of headcount ratio in Greece for the year 2013 and the corresponding sample sizes, mean square errors (MSE), standard errors (SD), coefficients of variance (CV) and confidence intervals

Nomos	Nomos_name	Sample_Size	EBLUP_FH	MSE	SD	LW95	UP95	CV
300001	Etolia and Akarnania	410	0.359588	0.000897	0.029953	0.299682	0.419493	8.329798
300003	Viotia	143	0.145612	0.00065	0.025498	0.094616	0.196608	17.51091
300004	Evia	341	0.257938	0.000776	0.027858	0.202221	0.313655	10.80044
300005	Evrytania	71	0.186228	0.001503	0.038764	0.1087	0.263755	20.81525
300006	Fthiotida	242	0.202991	0.000698	0.02642	0.150151	0.255831	13.01533
300007	Fokida	128	0.067338	0.000317	0.017795	0.031748	0.102928	26.42614
300011	Argolida	146	0.266861	0.001234	0.035134	0.196594	0.337128	13.16553
300012	Arkadia	206	0.195759	0.000791	0.028121	0.139517	0.252001	14.36511
300013	Achaia	794	0.264433	0.000349	0.018692	0.227048	0.301818	7.068857
300014	Ilia	242	0.219273	0.000762	0.027597	0.16408	0.274466	12.58553
300015	Korinthia	373	0.245251	0.000637	0.025247	0.194757	0.295745	10.29432
300016	Lakonia	63	0.196533	0.001471	0.038354	0.119825	0.27324	19.51518
300017	Messinia	333	0.238535	0.000822	0.028662	0.18121	0.29586	12.01601
300021	Zakynthos	106	0.287439	0.001581	0.039756	0.207927	0.36695	13.83102
300022	Kerkyra	113	0.213867	0.0012	0.03464	0.144587	0.283146	16.19691
300023	Kefallinia	69	0.263465	0.001746	0.041789	0.179888	0.347043	15.8612
300024	Lefkada	40	0.207018	0.00153	0.039112	0.128794	0.285242	18.89301
300031	Arta	126	0.247078	0.001304	0.03611	0.174858	0.319297	14.61479
300032	Thesprotia	31	0.269837	0.002241	0.047338	0.175162	0.364512	17.54303
300033	Loannina	417	0.235921	0.000647	0.025428	0.185065	0.286776	10.77812
300034	Preveza	178	0.224281	0.000929	0.030486	0.163308	0.285253	13.59297
300041	Karditsa	269	0.226347	0.000657	0.025633	0.175081	0.277612	11.32455
300042	Larissa	563	0.208337	0.000338	0.018397	0.171544	0.245131	8.830338
300043	Magnissia	430	0.229401	0.000588	0.024241	0.180918	0.277884	10.56733
300044	Trikala	83	0.218131	0.001213	0.034822	0.148488	0.287775	15.96371
300051	Grevena	37	0.205227	0.00225	0.047439	0.110349	0.300105	23.11537
300052	Drama	104	0.268494	0.001474	0.038388	0.191719	0.34527	14.29742
300053	Imathia	292	0.390111	0.001141	0.033775	0.32256	0.457661	8.657849
300054	Thessaloniki	1520	0.221772	0.000179	0.013384	0.195004	0.248541	6.035198
300055	Kavala	255	0.209479	0.000793	0.028157	0.153164	0.265794	13.44165
300056	Kastoria	135	0.236726	0.001267	0.035601	0.165525	0.307927	15.03878
300057	Kilkis	165	0.218162	0.000853	0.029204	0.159754	0.27657	13.38633
300058	Kozani	268	0.242886	0.000834	0.02888	0.185126	0.300646	11.89039
300059	Pella	199	0.318873	0.001193	0.034537	0.249799	0.387948	10.83103
300061	Pieria	370	0.301142	0.000732	0.027064	0.247014	0.355269	8.987099
300062	Serres	327	0.23905	0.00061	0.024702	0.189647	0.288453	10.33321
300063	Florina	106	0.282725	0.001349	0.036726	0.209273	0.356177	12.98998
300064	Chalkidiki and Aghion Oros	120	0.267997	0.001392	0.03731	0.193377	0.342616	13.92169
300071	Evros	332	0.213806	0.000573	0.023934	0.165937	0.261674	11.19442
300072	Xanthi	173	0.381644	0.001419	0.037671	0.306301	0.456986	9.870818
300073	Rodopi	230	0.328093	0.001497	0.038686	0.250721	0.405464	11.7911
300081	Dodekanissos	253	0.266095	0.000873	0.029552	0.206991	0.325198	11.10571
300082	Kyklades	173	0.211454	0.000827	0.028758	0.153938	0.26897	13.60011
300083	Lesvos	295	0.275406	0.000771	0.027775	0.219856	0.330956	10.08505
300084	Samos	28	0.177695	0.002498	0.049981	0.077733	0.277657	28.12731
300085	Chios	62	0.19987	0.001669	0.040857	0.118155	0.281585	20.44207
300091	Iraklio	622	0.213473	0.000353	0.018777	0.17592	0.251026	8.795755
300092	Lassithi	56	0.267276	0.001819	0.042647	0.181982	0.35257	15.95612
300093	Rethymno	125	0.300155	0.001593	0.039916	0.220322	0.379987	13.29858
300094	Chania	244	0.212673	0.000744	0.027272	0.158128	0.267217	12.82364
300101	Prefecture of Athens	3873	0.199659	0.000108	0.010408	0.178842	0.220476	5.213068
300102	Prefecture of East Attiki	871	0.22356	0.000314	0.017725	0.18811	0.259011	7.928594
300103	Prefecture of West Attiki	226	0.327091	0.001132	0.033649	0.259794	0.394389	10.28722
300104	Prefecture of Pireas	652	0.14582	0.000229	0.015119	0.115581	0.176058	10.36845

Table A8 EBLUP Fay and Herriot estimates of poverty gap in Greece for the year 2013 and the corresponding sample sizes, mean square errors (MSE), standard errors (SD), coefficients of variance (CV) and confidence intervals

Nomos	Nomos_name	Sample_Size	EBLUP_FH	MSE	SD	LW95	UP95	CV
300001	Etolia and Akarnania	410	0.133845	0.00017	0.01303	0.107786	0.159905	9.734895
300003	Viotia	143	0.01795	1.88E-05	0.004336	0.009273	0.026628	24.17004
300004	Evia	341	0.102224	0.000261	0.016155	0.069914	0.134534	15.80362
300005	Evrytania	71	0.045575	0.000234	0.015291	0.014993	0.076157	33.55083
300006	Fthiotida	242	0.080942	0.000189	0.013759	0.053424	0.10846	16.99851
300007	Fokida	128	0.007576	9.44E-06	0.003072	0.001433	0.01372	40.5451
300011	Argolida	146	0.096028	0.000276	0.016598	0.062832	0.129225	17.28488
300012	Arkadia	206	0.059681	0.000154	0.012414	0.034853	0.084509	20.8004
300013	Achaia	794	0.113811	0.000103	0.010166	0.09348	0.134142	8.931994
300014	Ilia	242	0.059589	8.48E-05	0.009209	0.041169	0.078009	15.45577
300015	Korinthia	373	0.097334	0.000154	0.012408	0.072517	0.122151	12.74827
300016	Lakonia	63	0.030258	7.30E-05	0.008544	0.013168	0.047348	28.24097
300017	Messinia	333	0.089818	0.000179	0.013376	0.063065	0.116571	14.89278
300021	Zakynthos	106	0.10862	0.00035	0.018701	0.071218	0.146023	17.21699
300022	Kerkyra	113	0.100616	0.000435	0.020862	0.058892	0.142341	20.73437
300023	Kefallinia	69	0.106685	0.000408	0.020204	0.066276	0.147093	18.9384
300024	Lefkada	40	0.025877	7.75E-05	0.008803	0.008272	0.043483	34.01724
300031	Arta	126	0.107932	0.000469	0.021659	0.064614	0.15125	20.06731
300032	Thesprotia	31	0.102687	0.000562	0.023717	0.055254	0.15012	23.0959
300033	Loannina	417	0.073705	9.45E-05	0.009721	0.05426	0.09315	13.19085
300034	Preveza	178	0.098892	0.000287	0.016942	0.065009	0.132775	17.13147
300041	Karditsa	269	0.093977	0.000195	0.013963	0.066052	0.121902	14.85746
300042	Larissa	563	0.097398	0.000111	0.010553	0.076292	0.118503	10.83494
300043	Magnissia	430	0.075136	7.93E-05	0.008905	0.057324	0.092949	11.85364
300044	Trikala	83	0.0683	0.000191	0.013818	0.040665	0.095936	20.23092
300051	Grevena	37	0.047657	0.000289	0.017006	0.013644	0.081669	35.68462
300052	Drama	104	0.116455	0.00038	0.019491	0.077473	0.155437	16.73682
300053	Imathia	292	0.176813	0.00038	0.019499	0.137816	0.215811	11.02795
300054	Thessaloniki	1520	0.081917	3.81E-05	0.006173	0.069574	0.094259	7.533475
300055	Kavala	255	0.100085	0.000286	0.016922	0.066241	0.13393	16.90786
300056	Kastoria	135	0.076947	0.000193	0.013903	0.049141	0.104753	18.06826
300057	Kilkis	165	0.078982	0.000165	0.012857	0.053268	0.104695	16.27825
300058	Kozani	268	0.074544	0.000112	0.010591	0.053362	0.095727	14.20815
300059	Pella	199	0.114287	0.000268	0.016364	0.081559	0.147016	14.31853
300061	Pieria	370	0.119966	0.000152	0.012342	0.095283	0.14465	10.28774
300062	Serres	327	0.082123	0.000121	0.011002	0.060119	0.104126	13.39667
300063	Florina	106	0.107881	0.000276	0.016615	0.07465	0.141112	15.40172
300064	Chalkidiki and Aghion Oros	120	0.067494	0.000169	0.012993	0.041509	0.093479	19.25009
300071	Evros	332	0.065266	9.54E-05	0.009767	0.04573	0.084803	14.96665
300072	Xanthi	173	0.167964	0.000414	0.020338	0.127288	0.20864	12.10864
300073	Rodopi	230	0.103838	0.000169	0.012986	0.077866	0.12981	12.50598
300081	Dodekanissos	253	0.07271	0.000114	0.010683	0.051343	0.094077	14.69335
300082	Kyklades	173	0.065706	0.000182	0.013494	0.038718	0.092693	20.5366
300083	Lesvos	295	0.101802	0.000143	0.011939	0.077923	0.125681	11.72813
300084	Samos	28	0.040975	0.000704	0.026532	-0.01209	0.094039	64.75229
300085	Chios	62	0.066315	0.00046	0.02144	0.023435	0.109196	32.33076
300091	Iraklio	622	0.076326	6.78E-05	0.008234	0.059854	0.092799	10.79106
300092	Lassithi	56	0.106632	0.000544	0.023324	0.059985	0.153279	21.87301
300093	Rethymno	125	0.125116	0.000444	0.021076	0.082964	0.167268	16.84521
300094	Chania	244	0.059196	9.40E-05	0.009695	0.039805	0.078587	16.37882
300101	Prefecture of Athens	3873	0.072545	1.51E-05	0.003886	0.064767	0.080323	5.360574
300102	Prefecture of East Attiki	871	0.074573	5.62E-05	0.007497	0.05958	0.089565	10.05237
300103	Prefecture of West Attiki	226	0.124792	0.000237	0.01538	0.094033	0.155551	12.32415
300104	Prefecture of Pireas	652	0.059283	6.02E-05	0.007759	0.043759	0.074806	13.09296

Table A9 EBLUP Fay and Herriot estimates of headcount ratio in Greece for the year 2009 and the corresponding sample sizes, mean square errors (MSE), standard errors (SD), coefficients of variance (CV) and confidence intervals

Nomos	Nomos_name	Sample_Size	EBLUP_FH	MSE	SD	LW95	UP95	CV
300001	Etolia and Akarnania	278	0.252298	0.001218	0.034906	0.182485	0.322111	13.83541
300003	Viotia	184	0.16709	0.000973	0.031196	0.104698	0.229483	18.67028
300004	Evia	254	0.353486	0.002076	0.045561	0.262363	0.444608	12.88913
300005	Evrytania	35	0.348094	0.005169	0.071895	0.204304	0.491884	20.65393
300006	Fthiotida	355	0.212686	0.001004	0.031689	0.149307	0.276065	14.8996
300007	Fokida	82	0.347905	0.004634	0.068077	0.211752	0.484058	19.5676
300011	Argolida	201	0.251062	0.002471	0.049708	0.151647	0.350478	19.7989
300012	Arkadia	171	0.279184	0.001693	0.041147	0.196889	0.361478	14.73836
300013	Achaia	573	0.256587	0.000873	0.029554	0.197478	0.315695	11.51819
300014	Ilia	369	0.35005	0.001264	0.035547	0.278956	0.421145	10.15484
300015	Korinthia	333	0.167513	0.000645	0.025405	0.116704	0.218323	15.16583
300016	Lakonia	123	0.272128	0.001882	0.04338	0.185369	0.358887	15.9409
300017	Messinia	220	0.332792	0.003897	0.062425	0.207942	0.457642	18.75798
300021	Zakynthos	31	0.289084	0.005476	0.074	0.141085	0.437083	25.59798
300022	Kerkyra	183	0.122294	0.000622	0.024936	0.072422	0.172167	20.39038
300024	Lefkada	28	0.292355	0.004146	0.06439	0.163574	0.421135	22.02465
300031	Arta	118	0.338786	0.003172	0.056319	0.226149	0.451424	16.62365
300032	Thesprotia	59	0.298093	0.00415	0.064418	0.169257	0.426928	21.61
300033	Loannina	204	0.228537	0.001677	0.040955	0.146627	0.310447	17.92054
300034	Preveza	106	0.343703	0.002988	0.054664	0.234375	0.453031	15.90445
300041	Karditsa	387	0.189802	0.000501	0.022384	0.145034	0.234569	11.79333
300042	Larissa	488	0.201386	0.000769	0.027737	0.145911	0.25686	13.77325
300043	Magnissia	346	0.15195	0.000515	0.022689	0.106571	0.197328	14.93213
300044	Trikala	127	0.257054	0.002045	0.045217	0.166619	0.347489	17.5906
300051	Grevena	67	0.366624	0.00411	0.064106	0.238412	0.494836	17.48547
300052	Drama	215	0.30804	0.002034	0.045099	0.217842	0.398237	14.64056
300053	Imathia	305	0.250638	0.001201	0.034653	0.181332	0.319945	13.82595
300054	Thessaloniki	1746	0.180621	0.000166	0.012899	0.154823	0.206418	7.141445
300055	Kavala	418	0.27356	0.000928	0.030465	0.212631	0.33449	11.13639
300056	Kastoria	98	0.282374	0.00285	0.053383	0.175609	0.38914	18.90494
300057	Kilkis	282	0.397259	0.002596	0.05095	0.295359	0.499158	12.82532
300058	Kozani	226	0.161344	0.000658	0.025656	0.110033	0.212656	15.90125
300059	Pella	261	0.269015	0.000925	0.030421	0.208174	0.329857	11.30819
300061	Pieria	315	0.313242	0.00131	0.036188	0.240867	0.385618	11.55262
300062	Serres	343	0.431188	0.001904	0.043639	0.343909	0.518467	10.12074
300063	Florina	147	0.334716	0.002761	0.052542	0.229632	0.4398	15.69742
300064	Chalkidiki and Aghion Oros	101	0.249685	0.002688	0.051847	0.145992	0.353378	20.76482
300071	Evros	246	0.266229	0.001536	0.039188	0.187854	0.344605	14.71958
300072	Xanthi	102	0.36039	0.003977	0.063062	0.234266	0.486513	17.49822
300073	Rodopi	226	0.316456	0.002475	0.049753	0.21695	0.415963	15.72198
300081	Dodekanissos	330	0.21663	0.001265	0.035566	0.145499	0.287762	16.41767
300082	Kyklades	225	0.118784	0.000626	0.025028	0.068728	0.16884	21.07006
300083	Lesvos	238	0.379511	0.003007	0.054834	0.269843	0.489178	14.44852
300084	Samos	22	0.295633	0.005816	0.076262	0.143108	0.448157	25.79626
300085	Chios	148	0.216001	0.001851	0.043019	0.129964	0.302038	19.91592
300091	Iraklio	570	0.1381	0.000453	0.021293	0.095515	0.180686	15.41841
300092	Lassithi	140	0.104125	0.000781	0.02794	0.048246	0.160005	26.8328
300093	Rethymno	71	0.048757	0.000392	0.019797	0.009163	0.08835	40.60301
300094	Chania	339	0.191149	0.000834	0.02888	0.13339	0.248908	15.10842
300101	Prefecture of Athens	3934	0.132662	5.86E-05	0.007654	0.117354	0.14797	5.769417
300102	Prefecture of East Attiki	680	0.109874	0.000226	0.015044	0.079786	0.139961	13.6918
300103	Prefecture of West Attiki	298	0.220794	0.001009	0.031761	0.157271	0.284317	14.38514
300104	Prefecture of Pireas	670	0.155255	0.000327	0.018089	0.119076	0.191434	11.65147

Table A10 EBLUP Fay and Herriot estimates of poverty gap in Greece for the year 2009 and the corresponding sample sizes, mean square errors (MSE), standard errors (SD), coefficients of variance (CV) and confidence intervals

Nomos	Nomos_name	Sample_Size	EBLUP_FH	MSE	SD	LW95	UP95	CV
300001	Etolia and Akarnania	278	0.059792	0.000106	0.010319	0.039153	0.080431	17.25893
300003	Viotia	184	0.061134	0.000199	0.014114	0.032905	0.089362	23.0874
300004	Evia	254	0.123624	0.000427	0.020671	0.082282	0.164966	16.72098
300005	Evrytania	35	0.103504	0.00055	0.023458	0.056589	0.15042	22.66353
300006	Fthiotida	355	0.047208	6.72E-05	0.008198	0.030813	0.063603	17.365
300007	Fokida	82	0.088581	0.000267	0.016328	0.055926	0.121236	18.43227
300011	Argolida	201	0.073154	0.000253	0.015916	0.041322	0.104986	21.75687
300012	Arkadia	171	0.103929	0.000376	0.01938	0.065169	0.142689	18.64745
300013	Achaia	573	0.072837	0.000113	0.010639	0.051559	0.094115	14.60633
300014	Ilia	369	0.112081	0.000186	0.013628	0.084825	0.139336	12.15901
300015	Korinthia	333	0.06835	0.000181	0.013464	0.041422	0.095278	19.69877
300016	Lakonia	123	0.080459	0.000253	0.015919	0.048621	0.112296	19.78482
300017	Messinia	220	0.104845	0.000601	0.02452	0.055806	0.153884	23.38655
300021	Zakynthos	31	0.088402	0.000418	0.02045	0.047502	0.129302	23.13299
300022	Kerkyra	183	0.037789	6.43E-05	0.008018	0.021753	0.053825	21.21746
300024	Lefkada	28	0.083751	0.000512	0.022626	0.038499	0.129004	27.01612
300031	Arta	118	0.089196	0.000363	0.01904	0.051117	0.127276	21.34606
300032	Thesprotia	59	0.061038	0.000251	0.015858	0.029321	0.092755	25.98111
300033	Loannina	204	0.036735	5.80E-05	0.007615	0.021506	0.051964	20.72847
300034	Preveza	106	0.099963	0.000366	0.019124	0.061715	0.138211	19.1311
300041	Karditsa	387	0.069663	0.000135	0.011605	0.046452	0.092873	16.65913
300042	Larissa	488	0.079067	0.000253	0.015907	0.047253	0.110882	20.11835
300043	Magnissia	346	0.042007	5.72E-05	0.007561	0.026884	0.05713	18.00028
300044	Trikala	127	0.089565	0.000458	0.021409	0.046748	0.132383	23.90289
300051	Grevena	67	0.107013	0.000502	0.022399	0.062215	0.151812	20.93116
300052	Drama	215	0.109476	0.000407	0.02018	0.069116	0.149836	18.43315
300053	Imathia	305	0.073056	0.000177	0.013322	0.046411	0.0997	18.23571
300054	Thessaloniki	1746	0.050667	2.02E-05	0.004494	0.04168	0.059655	8.869405
300055	Kavala	418	0.067942	0.000112	0.010606	0.04673	0.089153	15.61014
300056	Kastoria	98	0.106501	0.000545	0.023341	0.059819	0.153184	21.91647
300057	Kilkis	282	0.133789	0.000382	0.01954	0.094708	0.172869	14.60531
300058	Kozani	226	0.055355	0.000118	0.010841	0.033672	0.077038	19.58529
300059	Pella	261	0.085142	0.000188	0.013722	0.057699	0.112586	16.11598
300061	Pieria	315	0.076442	9.39E-05	0.009689	0.057063	0.095821	12.67558
300062	Serres	343	0.134017	0.000292	0.017098	0.099822	0.168212	12.75782
300063	Florina	147	0.097998	0.000532	0.023069	0.051859	0.144137	23.54081
300064	Chalkidiki and Aghion Oros	101	0.05079	0.000178	0.01336	0.024069	0.07751	26.3047
300071	Evros	246	0.07142	0.000137	0.011688	0.048045	0.094796	16.36485
300072	Xanthi	102	0.089232	0.000538	0.023192	0.042847	0.135616	25.99128
300073	Rodopi	226	0.090078	0.000389	0.019733	0.050612	0.129544	21.90676
300081	Dodekanissos	330	0.050459	7.37E-05	0.008583	0.033294	0.067624	17.00894
300082	Kyklades	225	0.02819	3.32E-05	0.005765	0.016661	0.03972	20.44922
300083	Lesvos	238	0.084202	0.000209	0.014454	0.055294	0.113109	17.16582
300084	Samos	22	0.072868	0.000491	0.022161	0.028547	0.11719	30.4119
300085	Chios	148	0.047467	0.000119	0.010911	0.025646	0.069289	22.9858
300091	Iraklio	570	0.041852	5.77E-05	0.007593	0.026666	0.057038	18.14246
300092	Lassithi	140	0.047017	0.000235	0.015313	0.01639	0.077644	32.56975
300093	Rethymno	71	0.006225	8.38E-06	0.002894	0.000436	0.012014	46.49811
300094	Chania	339	0.053772	0.000123	0.011095	0.031582	0.075962	20.63339
300101	Prefecture of Athens	3934	0.031811	5.31E-06	0.002305	0.027201	0.036421	7.245669
300102	Prefecture of East Attiki	680	0.043014	5.40E-05	0.007352	0.028311	0.057717	17.09094
300103	Prefecture of West Attiki	298	0.077895	0.000189	0.013759	0.050377	0.105413	17.66349
300104	Prefecture of Pireas	670	0.048257	4.91E-05	0.007007	0.034243	0.062272	14.52074

Table A11 Direct and EBLUP Fay and Herriot estimates of headcount ratio in Greece for the year 2013, their corresponding sample sizes, standard errors (SD), coefficients of variance (CV) and GIP1, GIP2 indicators

Nomos	Sample_Size	Direct_Headcount-ratio	SD_direct	CV_direct	EBLUP_FH	SD_FH	CV_FH	GIP1	GIP2
300001	410	0.401382	0.036449	9.080926	0.359588	0.029953	8.329798	1.216883	1.090174
300003	143	0.125327	0.028251	22.5417	0.145612	0.025498	17.51091	1.107963	1.287294
300004	341	0.277133	0.032941	11.88621	0.257938	0.027858	10.80044	1.18243	1.10053
300005	71	0.195504	0.052933	27.07525	0.186228	0.038764	20.81525	1.365536	1.300741
300006	242	0.197842	0.030702	15.51837	0.202991	0.02642	13.01533	1.162073	1.192315
300007	128	0.051831	0.01875	36.17431	0.067338	0.017795	26.42614	1.053646	1.368884
300011	146	0.307691	0.048982	15.91937	0.266861	0.035134	13.16553	1.394174	1.209171
300012	206	0.202858	0.032785	16.16178	0.195759	0.028121	14.36511	1.165871	1.125072
300013	794	0.272157	0.019959	7.333704	0.264433	0.018692	7.068857	1.067772	1.037467
300014	242	0.210363	0.032254	15.33247	0.219273	0.027597	12.58553	1.16876	1.218261
300015	373	0.251297	0.028902	11.50095	0.245251	0.025247	10.29432	1.144753	1.117213
300016	63	0.211938	0.055386	26.133	0.196533	0.038354	19.51518	1.444082	1.339111
300017	333	0.261845	0.03423	13.0725	0.238535	0.028662	12.01601	1.194237	1.087923
300021	106	0.280742	0.064222	22.87598	0.287439	0.039756	13.83102	1.615428	1.653962
300022	113	0.196137	0.046391	23.65214	0.213867	0.03464	16.19691	1.339231	1.460287
300023	69	0.380932	0.077898	20.44937	0.263465	0.041789	15.8612	1.864095	1.28927
300024	40	0.187066	0.062689	33.51147	0.207018	0.039112	18.89301	1.602799	1.773749
300031	126	0.260256	0.049419	18.98854	0.247078	0.03611	14.61479	1.368568	1.299269
300032	31	0.571288	0.17239	30.1756	0.269837	0.047338	17.54303	3.641713	1.720091
300033	417	0.246968	0.028888	11.69714	0.235921	0.025428	10.77812	1.136088	1.085267
300034	178	0.200535	0.037747	18.82332	0.224281	0.030486	13.59297	1.238173	1.384783
300041	269	0.215052	0.029087	13.52577	0.226347	0.025633	11.32455	1.134774	1.194376
300042	563	0.197767	0.019615	9.918352	0.208337	0.018397	8.830338	1.066225	1.123213
300043	430	0.227893	0.027369	12.00981	0.229401	0.024241	10.56733	1.129034	1.136504
300044	83	0.180892	0.04742	26.2144	0.218131	0.034822	15.96371	1.36178	1.642125
300051	37	0.270485	0.095891	35.45158	0.205227	0.047439	23.11537	2.021355	1.53368
300052	104	0.269402	0.056295	20.89617	0.268494	0.038388	14.29742	1.466477	1.461534
300053	292	0.463332	0.044141	9.526753	0.390111	0.033775	8.657849	1.30689	1.10036
300054	1520	0.221002	0.013733	6.213892	0.221772	0.013384	6.035198	1.026032	1.029609
300055	255	0.18601	0.033391	17.95107	0.209479	0.028157	13.44165	1.18586	1.335481
300056	135	0.249433	0.050008	20.04867	0.236726	0.035601	15.03878	1.404692	1.333131
300057	165	0.199317	0.035475	17.79833	0.218162	0.029204	13.38633	1.214739	1.32959
300058	268	0.242074	0.034979	14.44957	0.242886	0.02888	11.89039	1.211169	1.215231
300059	199	0.341688	0.046218	13.52628	0.318873	0.034537	10.83103	1.338196	1.248846
300061	370	0.302909	0.03146	10.38598	0.301142	0.027064	8.987099	1.162435	1.155654
300062	327	0.234048	0.027409	11.71104	0.23905	0.024702	10.33321	1.109627	1.13334
300063	106	0.272819	0.05278	19.34606	0.282725	0.036726	12.98998	1.437123	1.489307
300064	120	0.285653	0.054174	18.96499	0.267997	0.03731	13.92169	1.452013	1.362262
300071	332	0.210024	0.026885	12.80086	0.213806	0.023934	11.19442	1.12328	1.143504
300072	173	0.380403	0.04838	12.71815	0.381644	0.037671	9.870818	1.284273	1.28846
300073	230	0.355588	0.052988	14.90136	0.328093	0.038686	11.7911	1.369691	1.26378
300081	253	0.269381	0.035233	13.07939	0.266095	0.029552	11.10571	1.192263	1.177718
300082	173	0.191467	0.034609	18.07564	0.211454	0.028758	13.60011	1.203451	1.32908
300083	295	0.289127	0.033027	11.42304	0.275406	0.027775	10.08505	1.189101	1.132671
300084	28	0.443847	0.181398	40.86957	0.177695	0.049981	28.12731	3.629352	1.453021
300085	62	0.23003	0.067912	29.52323	0.19987	0.040857	20.44207	1.662172	1.444239
300091	622	0.198484	0.019958	10.05538	0.213473	0.018777	8.795755	1.062937	1.143209
300092	56	0.340881	0.08693	25.5016	0.267276	0.042647	15.95612	2.038368	1.598233
300093	125	0.284306	0.06058	21.30792	0.300155	0.039916	13.29858	1.517667	1.602271
300094	244	0.197354	0.031679	16.05168	0.212673	0.027272	12.82364	1.161563	1.251725
300101	3873	0.201803	0.010519	5.212367	0.199659	0.010408	5.213068	1.010603	0.999866
300102	871	0.223703	0.018537	8.286498	0.22356	0.017725	7.928594	1.045808	1.045141
300103	226	0.340047	0.042185	12.40557	0.327091	0.033649	10.28722	1.253686	1.20592
300104	652	0.141999	0.015676	11.03985	0.14582	0.015119	10.36845	1.036856	1.064754

Table A12 Direct and EBLUP Fay and Herriot estimates of poverty gap in Greece for the year 2013, their corresponding sample sizes, standard errors (SD), coefficients of variance (CV) and GIP1, GIP2 indicators

Nomos	Sample_Size	Direct_Poverty-gap	SD-direct	CV-direct	EBLUP_FH	SD_FH	CV_FH	GIP1	GIP2
300001	410	0.142915	0.014751	10.32139	0.133845	0.01303	9.734895	1.132092	1.060247
300003	143	0.017136	0.004387	25.60004	0.01795	0.004336	24.17004	1.011739	1.059164
300004	341	0.124678	0.020147	16.15914	0.102224	0.016155	15.80362	1.247099	1.022496
300005	71	0.045799	0.01778	38.82274	0.045575	0.015291	33.55083	1.162813	1.157132
300006	242	0.084714	0.015969	18.85103	0.080942	0.013759	16.99851	1.160654	1.108981
300007	128	0.007021	0.003089	44.00429	0.007576	0.003072	40.5451	1.005532	1.085317
300011	146	0.110253	0.021175	19.20599	0.096028	0.016598	17.28488	1.275734	1.111144
300012	206	0.064441	0.013804	21.4215	0.059681	0.012414	20.8004	1.111996	1.02986
300013	794	0.120094	0.010926	9.098143	0.113811	0.010166	8.931994	1.074835	1.018602
300014	242	0.057449	0.009748	16.96729	0.059589	0.009209	15.45577	1.058519	1.097797
300015	373	0.103305	0.013945	13.499	0.097334	0.012408	12.74827	1.123843	1.058889
300016	63	0.02844	0.008943	31.44705	0.030258	0.008544	28.24097	1.046756	1.113526
300017	333	0.101922	0.015312	15.02321	0.089818	0.013376	14.89278	1.144707	1.008758
300021	106	0.106228	0.025914	24.39443	0.10862	0.018701	17.21699	1.385674	1.416881
300022	113	0.114849	0.032774	28.53703	0.100616	0.020862	20.73437	1.570996	1.376315
300023	69	0.155102	0.031274	20.16345	0.106685	0.020204	18.9384	1.547881	1.064686
300024	40	0.019929	0.009289	46.61149	0.025877	0.008803	34.01724	1.055175	1.370231
300031	126	0.152025	0.035594	23.41304	0.107932	0.021659	20.06731	1.643364	1.166725
300032	31	0.189766	0.055207	29.09196	0.102687	0.023717	23.0959	2.327773	1.259615
300033	417	0.072885	0.010349	14.19947	0.073705	0.009721	13.19085	1.064617	1.076463
300034	178	0.100124	0.021761	21.73402	0.098892	0.016942	17.13147	1.284469	1.268661
300041	269	0.093209	0.016096	17.26851	0.093977	0.013963	14.85746	1.152779	1.162279
300042	563	0.096378	0.011427	11.85612	0.097398	0.010553	10.83494	1.08279	1.094249
300043	430	0.073229	0.009406	12.84404	0.075136	0.008905	11.85364	1.056198	1.083552
300044	83	0.058331	0.015995	27.42069	0.0683	0.013818	20.23092	1.157542	1.355385
300051	37	0.052448	0.020872	39.79517	0.047657	0.017006	35.68462	1.227319	1.115191
300052	104	0.121536	0.027469	22.6017	0.116455	0.019491	16.73682	1.409333	1.350418
300053	292	0.24384	0.027858	11.42459	0.176813	0.019499	11.02795	1.428685	1.035967
300054	1520	0.080789	0.006297	7.794411	0.081917	0.006173	7.533475	1.020168	1.034637
300055	255	0.098087	0.02162	22.04156	0.100085	0.016922	16.90786	1.277604	1.303628
300056	135	0.076452	0.016178	21.16152	0.076947	0.013903	18.06826	1.163671	1.171199
300057	165	0.075511	0.014591	19.32284	0.078982	0.012857	16.27825	1.134879	1.187034
300058	268	0.072067	0.011493	15.94775	0.074544	0.010591	14.20815	1.085135	1.122437
300059	199	0.113593	0.020357	17.92147	0.114287	0.016364	14.31853	1.244021	1.251628
300061	370	0.120276	0.01376	11.44045	0.119966	0.012342	10.28774	1.114919	1.112047
300062	327	0.077677	0.01185	15.2557	0.082123	0.011002	13.39667	1.077117	1.138767
300063	106	0.103116	0.020958	20.32482	0.107881	0.016615	15.40172	1.26136	1.319646
300064	120	0.062297	0.014702	23.59991	0.067494	0.012993	19.25009	1.131572	1.225964
300071	332	0.064245	0.010447	16.2614	0.065266	0.009767	14.96665	1.069603	1.086509
300072	173	0.180456	0.026992	14.95756	0.167964	0.020338	12.10864	1.327148	1.23528
300073	230	0.096135	0.01433	14.90574	0.103838	0.012986	12.50598	1.103469	1.191889
300081	253	0.069444	0.011505	16.56697	0.07271	0.010683	14.69335	1.076867	1.127515
300082	173	0.058281	0.015537	26.65856	0.065706	0.013494	20.5366	1.151412	1.2981
300083	295	0.105476	0.013294	12.60396	0.101802	0.011939	11.72813	1.11346	1.074677
300084	28	0.23757	0.123642	52.04453	0.040975	0.026532	64.75229	4.660105	0.803748
300085	62	0.091433	0.035431	38.75082	0.066315	0.02144	32.33076	1.652548	1.198574
300091	622	0.071539	0.008597	12.01654	0.076326	0.008234	10.79106	1.044021	1.113564
300092	56	0.180696	0.05374	29.74025	0.106632	0.023324	21.87301	2.304081	1.359678
300093	125	0.121115	0.032311	26.67824	0.125116	0.021076	16.84521	1.533086	1.583728
300094	244	0.054706	0.010315	18.85534	0.059196	0.009695	16.37882	1.063919	1.151203
300101	3873	0.072704	0.003907	5.374239	0.072545	0.003886	5.360574	1.005512	1.002549
300102	871	0.074739	0.00773	10.34217	0.074573	0.007497	10.05237	1.031082	1.028828
300103	226	0.128626	0.018056	14.03757	0.124792	0.01538	12.32415	1.174016	1.139029
300104	652	0.058523	0.008045	13.74734	0.059283	0.007759	13.09296	1.036932	1.04998

Table A13 Direct and EBLUP Fay and Herriot estimates of headcount ratio in Greece for the year 2009, their corresponding sample sizes, standard errors (SD), coefficients of variance (CV) and GIP1, GIP2 indicators

Nomos	Sample_Size	Direct_Headcount-ratio	SD_direct	CV_direct	EBLUP_FH	SD_FH	CV_FH	GIP1	GIP2
300001	278	0.238217	0.038245	16.05461	0.252298	0.034906	13.83541	1.095635	1.1604
300003	184	0.158318	0.033468	21.13958	0.16709	0.031196	18.67028	1.072814	1.132259
300004	254	0.397155	0.054367	13.68907	0.353486	0.045561	12.88913	1.19327	1.062064
300005	35	0.456709	0.145237	31.8007	0.348094	0.071895	20.65393	2.020121	1.539692
300006	355	0.205888	0.034212	16.61672	0.212686	0.031689	14.8996	1.079598	1.115246
300007	82	0.417272	0.111141	26.63507	0.347905	0.068077	19.5676	1.632582	1.361182
300011	201	0.27051	0.061572	22.76152	0.251062	0.049708	19.7989	1.238689	1.149636
300012	171	0.270788	0.0468	17.28295	0.279184	0.041147	14.73836	1.137385	1.172651
300013	573	0.264978	0.03147	11.87654	0.256587	0.029554	11.51819	1.064833	1.031112
300014	369	0.363437	0.039158	10.77432	0.35005	0.035547	10.15484	1.101578	1.061004
300015	333	0.164094	0.026592	16.20506	0.167513	0.025405	15.16583	1.046716	1.068524
300016	123	0.277055	0.050738	18.31336	0.272128	0.04338	15.9409	1.169627	1.148828
300017	220	0.47012	0.097606	20.76183	0.332792	0.062425	18.75798	1.563562	1.106827
300021	31	0.479406	0.14891	31.06146	0.289084	0.074	25.59798	2.012314	1.213434
300022	183	0.107219	0.026093	24.33622	0.122294	0.024936	20.39038	1.046392	1.193515
300024	28	0.255193	0.101933	39.9437	0.292355	0.06439	22.02465	1.583061	1.81359
300031	118	0.344103	0.075568	21.96074	0.338786	0.056319	16.62365	1.341786	1.321054
300032	59	0.288542	0.103311	35.80447	0.298093	0.064418	21.61	1.603761	1.656847
300033	204	0.219443	0.046701	21.2817	0.228537	0.040955	17.92054	1.140302	1.187559
300034	106	0.393241	0.072783	18.50854	0.343703	0.054664	15.90445	1.331465	1.163733
300041	387	0.178901	0.023158	12.94469	0.189802	0.022384	11.79333	1.034588	1.097627
300042	488	0.197933	0.029361	14.83369	0.201386	0.027737	13.77325	1.058527	1.076992
300043	346	0.144732	0.023528	16.25605	0.15195	0.022689	14.93213	1.036952	1.088662
300044	127	0.234552	0.053603	22.85331	0.257054	0.045217	17.5906	1.185447	1.299178
300051	67	0.417528	0.100017	23.95455	0.366624	0.064106	17.48547	1.560183	1.369969
300052	215	0.316476	0.053452	16.88983	0.30804	0.045099	14.64056	1.185228	1.153633
300053	305	0.251217	0.037978	15.11748	0.250638	0.034653	13.82595	1.095935	1.093413
300054	1746	0.182245	0.01303	7.149477	0.180621	0.012899	7.141445	1.010127	1.001125
300055	418	0.2765	0.032676	11.81771	0.27356	0.030465	11.13639	1.072583	1.06118
300056	98	0.302042	0.069861	23.12948	0.282374	0.053383	18.90494	1.308679	1.223462
300057	282	0.470629	0.064461	13.69669	0.397259	0.05095	12.82532	1.265181	1.067942
300058	226	0.1481	0.026852	18.13071	0.161344	0.025656	15.90125	1.04661	1.140206
300059	261	0.268012	0.032486	12.12105	0.269015	0.030421	11.30819	1.067883	1.071882
300061	315	0.328186	0.040119	12.22444	0.313242	0.036188	11.55262	1.108632	1.058152
300062	343	0.481144	0.050912	10.58152	0.431188	0.043639	10.12074	1.166661	1.045529
300063	147	0.374037	0.06793	18.16125	0.334716	0.052542	15.69742	1.292871	1.156958
300064	101	0.23681	0.066697	28.16474	0.249685	0.051847	20.76482	1.286426	1.356368
300071	246	0.261304	0.044221	16.92317	0.266229	0.039188	14.71958	1.128436	1.149705
300072	102	0.465969	0.094752	20.3343	0.36039	0.063062	17.49822	1.502519	1.162078
300073	226	0.326095	0.059319	18.19069	0.316456	0.049753	15.72198	1.192263	1.157022
300081	330	0.228706	0.039042	17.0708	0.21663	0.035566	16.41767	1.097742	1.039782
300082	225	0.10435	0.026199	25.10689	0.118784	0.025028	21.07006	1.0468	1.191591
300083	238	0.450904	0.072492	16.07694	0.379511	0.054834	14.44852	1.322027	1.112706
300084	22	0.380571	0.190228	49.98488	0.295633	0.076262	25.79626	2.494394	1.93768
300085	148	0.185181	0.047455	25.62617	0.216001	0.043019	19.91592	1.103126	1.286718
300091	570	0.134347	0.021951	16.33925	0.1381	0.021293	15.41841	1.030919	1.059724
300092	140	0.090286	0.029484	32.65671	0.104125	0.02794	26.8328	1.055288	1.217045
300093	71	0.038365	0.020345	53.0307	0.048757	0.019797	40.60301	1.027701	1.306078
300094	339	0.193932	0.030666	15.81254	0.191149	0.02888	15.10842	1.061843	1.046605
300101	3934	0.133447	0.007672	5.749034	0.132662	0.007654	5.769417	1.002361	0.996467
300102	680	0.109822	0.015249	13.88532	0.109874	0.015044	13.6918	1.013661	1.014134
300103	298	0.22279	0.034264	15.37971	0.220794	0.031761	14.38514	1.078804	1.069138
300104	670	0.154393	0.018466	11.96058	0.155255	0.018089	11.65147	1.020828	1.02653

Table A14 Direct and EBLUP Fay and Herriot estimates of poverty gap in Greece for the year 2009, their corresponding sample sizes, standard errors (SD), coefficients of variance (CV) and GIP1, GIP2 indicators

Nomos	Sample_Size	Direct_Poverty-gap	SD_direct	CV_direct	EBLUP_FH	SD_FH	CV_FH	GIP1	GIP2
300001	278	0.05752	0.010954	19.04442	0.059792	0.010319	17.25893	1.061521	1.103453
300003	184	0.060157	0.015906	26.44138	0.061134	0.014114	23.0874	1.126982	1.145273
300004	254	0.177562	0.028654	16.13757	0.123624	0.020671	16.72098	1.386194	0.965109
300005	35	0.11566	0.034344	29.69361	0.103504	0.023458	22.66353	1.464065	1.310194
300006	355	0.045058	0.008496	18.85643	0.047208	0.008198	17.365	1.036428	1.085887
300007	82	0.084115	0.018931	22.50582	0.088581	0.016328	18.43227	1.159446	1.221001
300011	201	0.075954	0.018677	24.58916	0.073154	0.015916	21.75687	1.173439	1.130179
300012	171	0.114397	0.024466	21.38712	0.103929	0.01938	18.64745	1.26245	1.14692
300013	573	0.072607	0.011283	15.53982	0.072837	0.010639	14.60633	1.060549	1.06391
300014	369	0.118334	0.01506	12.72694	0.112081	0.013628	12.15901	1.105105	1.046708
300015	333	0.07062	0.014911	21.11462	0.06835	0.013464	19.69877	1.107476	1.071875
300016	123	0.081342	0.018246	22.43069	0.080459	0.015919	19.78482	1.14618	1.133733
300017	220	0.155811	0.042493	27.27206	0.104845	0.02452	23.38655	1.733008	1.166143
300021	31	0.095775	0.027528	28.74213	0.088402	0.02045	23.13299	1.346101	1.242474
300022	183	0.033385	0.008275	24.78713	0.037789	0.008018	21.21746	1.032096	1.168242
300024	28	0.076687	0.032342	42.17408	0.083751	0.022626	27.01612	1.42939	1.561071
300031	118	0.091473	0.024286	26.55007	0.089196	0.01904	21.34606	1.27554	1.243793
300032	59	0.053528	0.018557	34.66852	0.061038	0.015858	25.98111	1.170197	1.334374
300033	204	0.033497	0.007857	23.45506	0.036735	0.007615	20.72847	1.031814	1.131538
300034	106	0.116287	0.024774	21.30423	0.099963	0.019124	19.1311	1.29544	1.113592
300041	387	0.065764	0.012488	18.9886	0.069663	0.011605	16.65913	1.076037	1.139831
300042	488	0.088395	0.018564	21.0012	0.079067	0.015907	20.11835	1.167034	1.043883
300043	346	0.040158	0.007795	19.40987	0.042007	0.007561	18.00028	1.03085	1.078309
300044	127	0.107087	0.030429	28.41496	0.089565	0.021409	23.90289	1.421329	1.188767
300051	67	0.113839	0.031626	27.78105	0.107013	0.022399	20.93116	1.411915	1.327258
300052	215	0.117326	0.025693	21.89859	0.109476	0.02018	18.43315	1.273186	1.188001
300053	305	0.074229	0.014785	19.91817	0.073056	0.013322	18.23571	1.109805	1.092262
300054	1746	0.050604	0.004541	8.972831	0.050667	0.004494	8.869405	1.010398	1.011661
300055	418	0.066857	0.011299	16.90032	0.067942	0.010606	15.61014	1.065374	1.08265
300056	98	0.108343	0.029464	27.19497	0.106501	0.023341	21.91647	1.26231	1.240847
300057	282	0.169787	0.025387	14.95206	0.133789	0.01954	14.60531	1.299202	1.023742
300058	226	0.051568	0.011556	22.40906	0.055355	0.010841	19.58529	1.065902	1.144178
300059	261	0.090622	0.015354	16.94293	0.085142	0.013722	16.11598	1.118972	1.051313
300061	315	0.078418	0.010199	13.00562	0.076442	0.009689	12.67558	1.052562	1.026037
300062	343	0.157912	0.020566	13.02374	0.134017	0.017098	12.75782	1.202858	1.020844
300063	147	0.134295	0.036722	27.34433	0.097998	0.023069	23.54081	1.591804	1.161572
300064	101	0.04756	0.014838	31.19854	0.05079	0.01336	26.3047	1.110637	1.186044
300071	246	0.072517	0.012602	17.37824	0.07142	0.011688	16.36485	1.078232	1.061924
300072	102	0.141702	0.035189	24.833	0.089232	0.023192	25.99128	1.517255	0.955436
300073	226	0.120702	0.025666	21.26397	0.090078	0.019733	21.90676	1.300652	0.970658
300081	330	0.049064	0.008855	18.04738	0.050459	0.008583	17.00894	1.031715	1.061052
300082	225	0.026676	0.005861	21.97196	0.02819	0.005765	20.44922	1.016738	1.074464
300083	238	0.085953	0.016339	19.00878	0.084202	0.014454	17.16582	1.130393	1.107362
300084	22	0.065275	0.032627	49.98488	0.072868	0.022161	30.4119	1.472317	1.643596
300085	148	0.042087	0.011665	27.71681	0.047467	0.010911	22.9858	1.06914	1.205823
300091	570	0.040799	0.007825	19.17839	0.041852	0.007593	18.14246	1.030515	1.0571
300092	140	0.035543	0.017686	49.76012	0.047017	0.015313	32.56975	1.15494	1.527802
300093	71	0.005642	0.002908	51.54146	0.006225	0.002894	46.49811	1.004591	1.108463
300094	339	0.052601	0.011883	22.59052	0.053772	0.011095	20.63339	1.071011	1.094852
300101	3934	0.031686	0.002312	7.295638	0.031811	0.002305	7.245669	1.002943	1.006896
300102	680	0.042888	0.007541	17.58225	0.043014	0.007352	17.09094	1.025736	1.028747
300103	298	0.086327	0.015201	17.60821	0.077895	0.013759	17.66349	1.104786	0.99687
300104	670	0.04746	0.007187	15.1442	0.048257	0.007007	14.52074	1.025713	1.042936

Table A15 Direct estimates of unemployment rate in Greece for the year 2013 and the corresponding sample sizes, variances (SD^2), standard errors (SD) and coefficients of variance (CV)

Nomos	Nomos_name	Sample_Size	Direct_unemployment	SD	SD^2	CV
300001	Etolia and Akarnania	148	0.366159	0.056131	0.003151	15.3297
300003	Viotia	57	0.257595	0.075007	0.005626	29.11827
300004	Evia	114	0.353285	0.061212	0.003747	17.32641
300005	Evrytania	34	0.418135	0.128252	0.016449	30.67241
300006	Fthiotida	82	0.229739	0.064375	0.004144	28.02094
300007	Fokida	54	0.344235	0.087456	0.007648	25.40579
300011	Argolida	63	0.391889	0.086657	0.007509	22.1126
300012	Arkadia	54	0.249615	0.079902	0.006384	32.00993
300013	Achaia	333	0.364097	0.037358	0.001396	10.26052
300014	Ilia	74	0.27852	0.075781	0.005743	27.20848
300015	Korinthia	139	0.230181	0.044715	0.001999	19.42593
300016	Lakonia	29	0.223749	0.097194	0.009447	43.43892
300017	Messinia	132	0.281452	0.051977	0.002702	18.46744
300021	Zakynthos	49	0.084478	0.056447	0.003186	66.81859
300022	Kerkyra	45	0.265878	0.082958	0.006882	31.20162
300023	Kefallinia	35	0.241303	0.092386	0.008535	38.28651
300031	Arta	49	0.315411	0.082928	0.006877	26.29197
300032	Thesprotia	11	0.057533	0.057437	0.003299	99.83189
300033	Loannina	165	0.353446	0.05563	0.003095	15.73934
300034	Preveza	71	0.244277	0.066867	0.004471	27.37354
300041	Karditsa	94	0.271683	0.058186	0.003386	21.41678
300042	Larissa	223	0.30808	0.039858	0.001589	12.93741
300043	Magnissia	159	0.265727	0.047149	0.002223	17.74327
300044	Trikala	43	0.342026	0.09739	0.009485	28.47453
300051	Grevena	9	0.359105	0.22179	0.049191	61.76181
300052	Drama	32	0.424288	0.119486	0.014277	28.16158
300053	Imathia	92	0.369234	0.069092	0.004774	18.71224
300054	Thessaloniki	643	0.354214	0.027201	0.00074	7.679327
300055	Kavala	107	0.251399	0.053428	0.002855	21.25216
300056	Kastoria	43	0.256626	0.087443	0.007646	34.0742
300057	Kilkis	58	0.362807	0.082634	0.006828	22.77636
300058	Kozani	102	0.263487	0.056849	0.003232	21.57563
300059	Pella	77	0.35696	0.073409	0.005389	20.56501
300061	Pieria	161	0.290534	0.045001	0.002025	15.48919
300062	Serres	103	0.235838	0.049666	0.002467	21.05935
300063	Florina	34	0.284075	0.093371	0.008718	32.86854
300064	Chalkidiki and Aghion Oros	42	0.31714	0.096979	0.009405	30.57914
300071	Evros	117	0.175833	0.041006	0.001682	23.32125
300072	Xanthi	62	0.367792	0.080159	0.006426	21.79474
300073	Rodopi	92	0.253151	0.070636	0.004989	27.90279
300081	Dodekanissos	95	0.204638	0.048968	0.002398	23.92915
300082	Kyklades	70	0.270393	0.066745	0.004455	24.68446
300083	Lesvos	111	0.269242	0.05423	0.002941	20.14168
300084	Samos	11	0.237297	0.106149	0.011268	44.7326
300085	Chios	20	0.29724	0.127209	0.016182	42.79671
300091	Iraklio	257	0.1993	0.029406	0.000865	14.7545
300092	Lassithi	25	0.197066	0.104275	0.010873	52.91396
300093	Rethymno	41	0.187678	0.082384	0.006787	43.89649
300094	Chania	101	0.287859	0.054728	0.002995	19.01191
300101	Prefecture of Athens	1672	0.301099	0.017975	0.000323	5.969954
300102	Prefecture of East Attiki	381	0.248986	0.027847	0.000775	11.18434
300103	Prefecture of West Attiki	107	0.457323	0.070573	0.004981	15.43183
300104	Prefecture of Pireas	315	0.326289	0.036375	0.001323	11.14801

Table A16 Direct estimates of unemployment rate in Greece for the year 2009 and the corresponding sample sizes, variances (SD^2), standard errors (SD) and coefficients of variance (CV)

Nomos	Nomos_name	Sample_Size	Direct_Unemployment	SD	SD^2	CV
300001	Etolia and Akarnania	108	0.086476	0.035262	0.001243	40.77658
300003	Viotia	88	0.062329	0.025053	0.000628	40.1951
300004	Evia	93	0.180645	0.053715	0.002885	29.73535
300005	Evrytania	6	0.365171	0.298974	0.089386	81.87227
300006	Fthiotida	134	0.044861	0.01989	0.000396	44.33742
300007	Fokida	36	0.276827	0.140646	0.019781	50.80637
300011	Argolida	78	0.026904	0.023907	0.000572	88.85889
300012	Arkadia	62	0.006372	0.005457	2.98E-05	85.64908
300013	Achaia	218	0.157377	0.04322	0.001868	27.46249
300014	Ilia	147	0.136036	0.040811	0.001666	30.00017
300015	Korinthia	146	0.112331	0.034762	0.001208	30.94597
300016	Lakonia	59	0.028164	0.021978	0.000483	78.03425
300017	Messinia	66	0.037545	0.018824	0.000354	50.13595
300021	Zakynthos	12	0.25493	0.180193	0.03247	70.68352
300022	Kerkyra	78	0.017624	0.017305	0.000299	98.18893
300023	Kefallinia	10	0.098876	0.098829	0.009767	99.95334
300032	Thesprotia	31	0.142828	0.094659	0.00896	66.27483
300033	Loannina	82	0.120861	0.052247	0.00273	43.22902
300034	Preveza	47	0.200089	0.063513	0.004034	31.7424
300041	Karditsa	157	0.101197	0.03428	0.001175	33.87402
300042	Larissa	197	0.150466	0.043159	0.001863	28.68388
300043	Magnissia	154	0.116818	0.034202	0.00117	29.27818
300044	Trikala	48	0.071326	0.05521	0.003048	77.40477
300051	Grevena	28	0.039759	0.039734	0.001579	99.93734
300052	Drama	72	0.067455	0.035325	0.001248	52.36765
300053	Imathia	123	0.12481	0.041197	0.001697	33.008
300054	Thessaloniki	755	0.105184	0.01493	0.000223	14.19429
300055	Kavala	165	0.114729	0.030218	0.000913	26.33891
300056	Kastoria	39	0.077433	0.064228	0.004125	82.9469
300057	Kilkis	110	0.208489	0.062421	0.003896	29.93959
300058	Kozani	91	0.142294	0.057382	0.003293	40.3263
300059	Pella	100	0.075344	0.029511	0.000871	39.16753
300061	Pieria	128	0.066224	0.027683	0.000766	41.80274
300062	Serres	137	0.221002	0.059082	0.003491	26.73349
300063	Florina	59	0.143299	0.065068	0.004234	45.40751
300064	Chalkidiki and Aghion Oros	42	0.106121	0.068854	0.004741	64.88278
300071	Evros	89	0.091101	0.042185	0.00178	46.30615
300072	Xanthi	43	0.148178	0.078723	0.006197	53.12722
300073	Rodopi	94	0.012347	0.006693	4.48E-05	54.20794
300081	Dodekanissos	161	0.125046	0.03698	0.001367	29.57281
300082	Kyklades	97	0.065402	0.032926	0.001084	50.34365
300083	Lesvos	85	0.098959	0.052716	0.002779	53.27014
300085	Chios	54	0.113969	0.067066	0.004498	58.84609
300091	Iraklio	273	0.051495	0.017863	0.000319	34.68775
300092	Lassithi	63	0.028079	0.028069	0.000788	99.9641
300093	Rethymno	27	0.046932	0.02859	0.000817	60.91936
300094	Chania	162	0.172964	0.043172	0.001864	24.96026
300101	Prefecture of Athens	1729	0.101941	0.009813	9.63E-05	9.626502
300102	Prefecture of East Attiki	294	0.068738	0.017647	0.000311	25.67207
300103	Prefecture of West Attiki	102	0.095405	0.03752	0.001408	39.32729
300104	Prefecture of Pireas	271	0.067866	0.017605	0.00031	25.94065

Table A17 EBLUP Fay and Herriot estimates of unemployment rate in Greece for the year 2013 and the corresponding sample sizes, mean square errors (MSE), standard errors (SD), coefficients of variance (CV) and confidence intervals

Nomos	**Nomos_name**	**Sample_size**	**EBLUP-FH**	**MSE**	**SD**	**LW95**	**UP95**	**CV**
300001	Etolia and Akarnania	148	0.349249	0.000681	0.026101	0.297047	0.40145	7.473379
300003	Viotia	57	0.261554	0.000355	0.018846	0.223862	0.299247	7.205549
300004	Evia	114	0.330374	0.00041	0.020239	0.289897	0.370852	6.126066
300005	Evrytania	34	0.417948	0.002633	0.051317	0.315313	0.520583	12.27842
300006	Fthiotida	82	0.326649	0.000468	0.021641	0.283368	0.369931	6.625048
300007	Fokida	54	0.36969	0.00075	0.027395	0.3149	0.424481	7.410283
300011	Argolida	63	0.226638	0.000651	0.025518	0.175602	0.277673	11.2592
300012	Arkadia	54	0.260303	0.0006	0.024499	0.211306	0.3093	9.411543
300013	Achaia	333	0.351014	0.000721	0.026852	0.29731	0.404718	7.649827
300014	Ilia	74	0.331501	0.000424	0.020597	0.290308	0.372694	6.213109
300015	Korinthia	139	0.250109	0.00043	0.020732	0.208645	0.291574	8.289294
300016	Lakonia	29	0.231809	0.000739	0.027185	0.177439	0.28618	11.72739
300017	Messinia	132	0.261616	0.000403	0.020063	0.221491	0.301742	7.668775
300021	Zakynthos	49	0.192977	0.000491	0.022149	0.148679	0.237274	11.47733
300022	Kerkyra	45	0.233138	0.000446	0.021115	0.190908	0.275369	9.056955
300023	Kefallinia	35	0.25199	0.000387	0.01967	0.212649	0.29133	7.806012
300031	Arta	49	0.256378	0.000583	0.024143	0.208091	0.304664	9.417019
300032	Thesprotia	11	0.147701	0.000745	0.027303	0.093096	0.202306	18.48507
300033	Loannina	165	0.270632	0.000504	0.022451	0.22573	0.315533	8.295674
300034	Preveza	71	0.233068	0.000456	0.021349	0.190369	0.275767	9.160144
300041	Karditsa	94	0.277265	0.000498	0.022327	0.232612	0.321919	8.052457
300042	Larissa	223	0.265695	0.00053	0.023014	0.219666	0.311724	8.661976
300043	Magnissia	159	0.313087	0.000394	0.019858	0.27337	0.352804	6.342806
300044	Trikala	43	0.249932	0.000489	0.022104	0.205724	0.294139	8.843882
300051	Grevena	9	0.285591	0.000635	0.02519	0.235212	0.335971	8.820219
300052	Drama	32	0.330115	0.000474	0.021777	0.28656	0.37367	6.596903
300053	Imathia	92	0.278634	0.000335	0.018302	0.242029	0.315238	6.568538
300054	Thessaloniki	643	0.336468	0.000536	0.023146	0.290176	0.38276	6.879104
300055	Kavala	107	0.278669	0.00047	0.021679	0.23531	0.322028	7.779655
300056	Kastoria	43	0.212912	0.000478	0.021864	0.169184	0.25664	10.26904
300057	Kilkis	58	0.325915	0.000565	0.023771	0.278372	0.373458	7.293751
300058	Kozani	102	0.316332	0.000665	0.025778	0.264775	0.367888	8.149062
300059	Pella	77	0.258706	0.000459	0.02143	0.215847	0.301565	8.283374
300061	Pieria	161	0.301058	0.00045	0.021208	0.258642	0.343473	7.044373
300062	Serres	103	0.272276	0.000622	0.024935	0.222405	0.322147	9.158159
300063	Florina	34	0.27329	0.000344	0.018554	0.236182	0.310399	6.789248
300064	Chalkidiki and Aghion Oros	42	0.267005	0.000724	0.026909	0.213187	0.320824	10.07816
300071	Evros	117	0.208848	0.000791	0.028118	0.152611	0.265084	13.46362
300072	Xanthi	62	0.370802	0.00187	0.043245	0.284312	0.457293	11.66259
300073	Rodopi	92	0.216402	0.000518	0.022766	0.17087	0.261934	10.52029
300081	Dodekanissos	95	0.212777	0.000632	0.025131	0.162514	0.263039	11.81106
300082	Kyklades	70	0.218157	0.00062	0.024899	0.168359	0.267954	11.41323
300083	Lesvos	111	0.258009	0.000435	0.020852	0.216304	0.299714	8.082063
300084	Samos	11	0.237389	0.00042	0.020503	0.196382	0.278396	8.637083
300085	Chios	20	0.268133	0.000466	0.021577	0.224979	0.311287	8.047063
300091	Iraklio	257	0.222529	0.000568	0.023829	0.17487	0.270187	10.7084
300092	Lassithi	25	0.133956	0.000798	0.028254	0.077448	0.190464	21.09212
300093	Rethymno	41	0.252498	0.001092	0.03305	0.186399	0.318598	13.08911
300094	Chania	101	0.256135	0.000382	0.019543	0.217049	0.295221	7.62994
300101	Prefecture of Athens	1672	0.289484	0.000585	0.024186	0.241112	0.337856	8.354827
300102	Prefecture of East Attiki	381	0.270181	0.000593	0.024356	0.221469	0.318893	9.014732
300103	Prefecture of West Attiki	107	0.363625	0.000827	0.028752	0.306121	0.421128	7.906958
300104	Prefecture of Pireas	315	0.384775	0.000644	0.025373	0.33403	0.43552	6.594134

Table A18 EBLUP Fay and Herriot estimates of unemployment rate in Greece for the year 2009 and the corresponding sample sizes, mean square errors (MSE), standard errors (SD), coefficients of variance (CV) and confidence intervals

Nomos	Nomos_name	Sample_Size	EBLUP-FH	MSE	SD	LW95	UP95	CV
300001	Etolia and Akarnania	108	0.091531	0.000627	0.025042	0.041447	0.141615	27.35886
300003	Viotia	88	0.061632	0.000432	0.020791	0.02005	0.103213	33.73395
300004	Evia	93	0.122551	0.000854	0.029227	0.064096	0.181006	23.84915
300005	Evrytania	6	0.020125	0.001454	0.038137	-0.05615	0.0964	189.5007
300006	Fthiotida	134	0.0535	0.000307	0.01752	0.018461	0.08854	32.74734
300007	Fokida	36	0.090423	0.001287	0.035868	0.018687	0.162159	39.66713
300011	Argolida	78	0.037446	0.000407	0.020177	-0.00291	0.077799	53.88276
300012	Arkadia	62	0.00812	2.94E-05	0.005422	-0.00272	0.018962	66.76316
300013	Achaia	218	0.129783	0.000767	0.027687	0.074409	0.185157	21.33317
300014	Ilia	147	0.111368	0.000713	0.026708	0.057952	0.164783	23.98178
300015	Korinthia	146	0.081332	0.000629	0.02508	0.031171	0.131493	30.83693
300016	Lakonia	59	0.028156	0.00038	0.019504	-0.01085	0.067165	69.2722
300017	Messinia	66	0.042401	0.000286	0.016926	0.008548	0.076253	39.91952
300021	Zakynthos	12	0.051923	0.00119	0.0345	-0.01708	0.120923	66.44549
300022	Kerkyra	78	0.030115	0.000247	0.015704	-0.00129	0.061522	52.14498
300023	Kefallinia	10	0.080437	0.001081	0.032877	0.014683	0.146191	40.87294
300032	Thesprotia	31	0.076602	0.000994	0.031523	0.013556	0.139648	41.15194
300033	Loannina	82	0.104813	0.000852	0.029182	0.046449	0.163178	27.84224
300034	Preveza	47	0.103119	0.000883	0.029714	0.043691	0.162547	28.8152
300041	Karditsa	157	0.076757	0.000622	0.024938	0.026881	0.126633	32.48943
300042	Larissa	197	0.10965	0.000726	0.026952	0.055747	0.163553	24.57962
300043	Magnissia	154	0.111658	0.000618	0.024868	0.061922	0.161394	22.27144
300044	Trikala	48	0.075285	0.000846	0.029092	0.017102	0.133469	38.6419
300051	Grevena	28	0.046924	0.00076	0.027561	-0.0082	0.102047	58.73633
300052	Drama	72	0.079665	0.000636	0.025211	0.029242	0.130087	31.64666
300053	Imathia	123	0.100193	0.000703	0.026507	0.04718	0.153206	26.45548
300054	Thessaloniki	755	0.103859	0.000195	0.013969	0.075921	0.131797	13.44995
300055	Kavala	165	0.099696	0.00053	0.023029	0.053638	0.145755	23.09942
300056	Kastoria	39	0.082493	0.000884	0.029728	0.023038	0.141949	36.03666
300057	Kilkis	110	0.110317	0.000903	0.030048	0.050222	0.170413	27.23774
300058	Kozani	91	0.128324	0.00097	0.031144	0.066036	0.190611	24.26969
300059	Pella	100	0.07435	0.000516	0.022717	0.028917	0.119783	30.55376
300061	Pieria	128	0.080788	0.000484	0.022009	0.03677	0.124805	27.24291
300062	Serres	137	0.109712	0.000881	0.029683	0.050346	0.169079	27.05557
300063	Florina	59	0.109886	0.000931	0.030506	0.048875	0.170898	27.76131
300064	Chalkidiki and Aghion Oros	42	0.103436	0.000934	0.030559	0.042319	0.164554	29.54353
300071	Evros	89	0.096548	0.000742	0.027244	0.042061	0.151036	28.21784
300072	Xanthi	43	0.122072	0.001055	0.03248	0.057112	0.187033	26.60736
300073	Rodopi	94	0.015075	4.36E-05	0.006603	0.001868	0.028281	43.8042
300081	Dodekanissos	161	0.113701	0.000698	0.026416	0.060869	0.166532	23.23277
300082	Kyklades	97	0.073602	0.000576	0.024008	0.025587	0.121617	32.6181
300083	Lesvos	85	0.090554	0.000869	0.029479	0.031596	0.149512	32.55391
300085	Chios	54	0.142093	0.001392	0.037306	0.067481	0.216705	26.25462
300091	Iraklio	273	0.056801	0.000264	0.01624	0.024322	0.089281	28.59017
300092	Lassithi	63	0.02654	0.000542	0.023287	-0.02003	0.073115	87.74437
300093	Rethymno	27	0.059662	0.000501	0.022388	0.014886	0.104438	37.52458
300094	Chania	162	0.111249	0.000731	0.027032	0.057185	0.165313	24.29846
300101	Prefecture of Athens	1729	0.099196	9.12E-05	0.00955	0.080094	0.118299	9.628451
300102	Prefecture of East Attiki	294	0.07514	0.00026	0.016136	0.042867	0.107412	21.47492
300103	Prefecture of West Attiki	102	0.110881	0.000719	0.026819	0.057243	0.16452	24.18731
300104	Prefecture of Pireas	271	0.07823	0.000258	0.016062	0.046105	0.110355	20.53227

Table A19 Direct and EBLUP Fay and Herriot estimates of unemployment rate in Greece for the year 2013, their corresponding sample sizes, standard errors (SD), coefficients of variance (CV) and GIP1, GIP2 indicators

Nomos	Sample_size	Direct_unemployment	SD_direct	cv_direct	EBLUP -FH	SD_FH	CV_FH	GIP1	GIP2
300001	148	0.366159	0.056131	15.3297	0.349249	0.026101	7.473379	2.150562	2.051242
300003	57	0.257595	0.075007	29.11827	0.261554	0.018846	7.205549	3.979912	4.04109
300004	114	0.353285	0.061212	17.32641	0.330374	0.020239	6.126066	3.024447	2.828309
300005	34	0.418135	0.128252	30.67241	0.417948	0.051317	12.27842	2.499191	2.498075
300006	82	0.229739	0.064375	28.02094	0.326649	0.021641	6.625048	2.974725	4.229545
300007	54	0.344235	0.087456	25.40579	0.36969	0.027395	7.410283	3.192381	3.42845
300011	63	0.391889	0.086657	22.1126	0.226638	0.025518	11.2592	3.395966	1.963959
300012	54	0.249615	0.079902	32.00993	0.260303	0.024499	9.411543	3.261484	3.401135
300013	333	0.364097	0.037358	10.26052	0.351014	0.026852	7.649827	1.391266	1.341274
300014	74	0.27852	0.075781	27.20848	0.331501	0.020597	6.213109	3.679306	4.379205
300015	139	0.230181	0.044715	19.42593	0.250109	0.020732	8.289294	2.156774	2.343497
300016	29	0.223749	0.097194	43.43892	0.231809	0.027185	11.72739	3.575258	3.704057
300017	132	0.281452	0.051977	18.46744	0.261616	0.020063	7.668775	2.590723	2.408135
300021	49	0.084478	0.056447	66.81859	0.192977	0.022149	11.47733	2.548562	5.821789
300022	45	0.265878	0.082958	31.20162	0.233138	0.021115	9.056955	3.92883	3.445045
300023	35	0.241303	0.092386	38.28651	0.25199	0.01967	7.806012	4.696738	4.904746
300031	49	0.315411	0.082928	26.29197	0.256378	0.024143	9.417019	3.434836	2.791963
300032	11	0.057533	0.057437	99.83189	0.147701	0.027303	18.48507	2.103701	5.400676
300033	165	0.353446	0.05563	15.73934	0.270632	0.022451	8.295674	2.477872	1.897295
300034	71	0.244277	0.066867	27.37354	0.233068	0.021349	9.160144	3.132049	2.988331
300041	94	0.271683	0.058186	21.41678	0.277265	0.022327	8.052457	2.606112	2.659658
300042	223	0.30808	0.039858	12.93741	0.265695	0.023014	8.661976	1.73185	1.493587
300043	159	0.265727	0.047149	17.74327	0.313087	0.019858	6.342806	2.374237	2.797386
300044	43	0.342026	0.09739	28.47453	0.249932	0.022104	8.843882	4.406077	3.219687
300051	9	0.359105	0.22179	61.76181	0.285591	0.02519	8.820219	8.804767	7.002299
300052	32	0.424288	0.119486	28.16158	0.330115	0.021777	6.596903	5.486705	4.268908
300053	92	0.369234	0.069092	18.71224	0.278634	0.018302	6.568538	3.775072	2.848768
300054	643	0.354214	0.027201	7.679327	0.336468	0.023146	6.879104	1.175204	1.116327
300055	107	0.251399	0.053428	21.25216	0.278669	0.021679	7.779655	2.464438	2.731761
300056	43	0.256626	0.087443	34.0742	0.212912	0.021864	10.26904	3.999402	3.31815
300057	58	0.362807	0.082634	22.77636	0.325915	0.023771	7.293751	3.476198	3.122723
300058	102	0.263487	0.056849	21.57563	0.316332	0.025778	8.149062	2.205325	2.647621
300059	77	0.35696	0.073409	20.56501	0.258706	0.02143	8.283374	3.425584	2.482686
300061	161	0.290534	0.045001	15.48919	0.301058	0.021208	7.044373	2.121941	2.198803
300062	103	0.235838	0.049666	21.05935	0.272276	0.024935	9.158159	1.991777	2.299517
300063	34	0.284075	0.093371	32.86854	0.27329	0.018554	6.789248	5.032313	4.841264
300064	42	0.31714	0.096979	30.57914	0.267005	0.026909	10.07816	3.603922	3.034198
300071	117	0.175833	0.041006	23.32125	0.208848	0.028118	13.46362	1.458344	1.732168
300072	62	0.367792	0.080159	21.79474	0.370802	0.043245	11.66259	1.853599	1.868774
300073	92	0.253151	0.070636	27.90279	0.216402	0.022766	10.52029	3.102689	2.652283
300081	95	0.204638	0.048968	23.92915	0.212777	0.025131	11.81106	1.948503	2.025996
300082	70	0.270393	0.066745	24.68446	0.218157	0.024899	11.41323	2.680655	2.162793
300083	111	0.269242	0.05423	20.14168	0.258009	0.020852	8.082063	2.600646	2.492146
300084	11	0.237297	0.106149	44.7326	0.237389	0.020503	8.637083	5.177133	5.179133
300085	20	0.29724	0.127209	42.79671	0.268133	0.021577	8.047063	5.895625	5.318302
300091	257	0.1993	0.029406	14.7545	0.222529	0.023829	10.7084	1.234014	1.377844
300092	25	0.197066	0.104275	52.91396	0.133956	0.028254	21.09212	3.69062	2.508707
300093	41	0.187678	0.082384	43.89649	0.252498	0.03305	13.08911	2.492728	3.353664
300094	101	0.287859	0.054728	19.01191	0.256135	0.019543	7.62994	2.800374	2.491751
300101	1672	0.301099	0.017975	5.969954	0.289484	0.024186	8.354827	0.743222	0.714551
300102	381	0.248986	0.027847	11.18434	0.270181	0.024356	9.014732	1.143348	1.240673
300103	107	0.457323	0.070573	15.43183	0.363625	0.028752	7.906958	2.454582	1.951677
300104	315	0.326289	0.036375	11.14801	0.384775	0.025373	6.594134	1.433624	1.690594

Table A20 Direct and EBLUP Fay and Herriot estimates of unemployment rate in Greece for the year 2009, their corresponding sample sizes, standard errors (SD), coefficients of variance (CV) and GIP1, GIP2 indicators

Nomos	Sample_Size	Direct_Unemployment	SD_direct	CV_direct	EBLUP_FH	SD_FH	CV_FH	GIP1	GIP2
300001	108	0.086476	0.035262	40.77658	0.091531	0.025042	27.35886	1.408114	1.490434
300003	88	0.062329	0.025053	40.1951	0.061632	0.020791	33.73395	1.20501	1.191533
300004	93	0.180645	0.053715	29.73535	0.122551	0.029227	23.84915	1.837845	1.24681
300005	6	0.365171	0.298974	81.87227	0.020125	0.038137	189.5007	7.839392	0.432042
300006	134	0.044861	0.01989	44.33742	0.0535	0.01752	32.74734	1.135288	1.353924
300007	36	0.276827	0.140646	50.80637	0.090423	0.035868	39.66713	3.92119	1.280818
300011	78	0.026904	0.023907	88.85889	0.037446	0.020177	53.88276	1.184871	1.649115
300012	62	0.006372	0.005457	85.64908	0.00812	0.005422	66.76316	1.006476	1.282879
300013	218	0.157377	0.04322	27.46249	0.129783	0.027687	21.33317	1.56102	1.287314
300014	147	0.136036	0.040811	30.00017	0.111368	0.026708	23.98178	1.528053	1.250957
300015	146	0.112331	0.034762	30.94597	0.081332	0.02508	30.83693	1.386019	1.003536
300016	59	0.028164	0.021978	78.03425	0.028156	0.019504	69.2722	1.126813	1.126487
300017	66	0.037545	0.018824	50.13595	0.042401	0.016926	39.91952	1.112109	1.255926
300021	12	0.25493	0.180193	70.68352	0.051923	0.0345	66.44549	5.222955	1.063782
300022	78	0.017624	0.017305	98.18893	0.030115	0.015704	52.14498	1.101962	1.882999
300023	10	0.098876	0.098829	99.95334	0.080437	0.032877	40.87294	3.006032	2.445465
300032	31	0.142828	0.094659	66.27483	0.076602	0.031523	41.15194	3.002836	1.610491
300033	82	0.120861	0.052247	43.22902	0.104813	0.029182	27.84224	1.790354	1.552641
300034	47	0.200089	0.063513	31.7424	0.103119	0.029714	28.8152	2.137485	1.101585
300041	157	0.101197	0.03428	33.87402	0.076757	0.024938	32.48943	1.374593	1.042617
300042	197	0.150466	0.043159	28.68388	0.10965	0.026952	24.57962	1.601366	1.166978
300043	154	0.116818	0.034202	29.27818	0.111658	0.024868	22.27144	1.375355	1.314607
300044	48	0.071326	0.05521	77.40477	0.075285	0.029092	38.6419	1.897779	2.00313
300051	28	0.039759	0.039734	99.93734	0.046924	0.027561	58.73633	1.441647	1.701457
300052	72	0.067455	0.035325	52.36765	0.079665	0.025211	31.64666	1.401152	1.654761
300053	123	0.12481	0.041197	33.008	0.100193	0.026507	26.45548	1.554238	1.247681
300054	755	0.105184	0.01493	14.19429	0.103859	0.013969	13.44995	1.068803	1.055341
300055	165	0.114729	0.030218	26.33891	0.099696	0.023029	23.09942	1.312175	1.140241
300056	39	0.077433	0.064228	82.9469	0.082493	0.029728	36.03666	2.160533	2.301737
300057	110	0.208489	0.062421	29.93959	0.110317	0.030048	27.23774	2.077371	1.099195
300058	91	0.142294	0.057382	40.3263	0.128324	0.031144	24.26969	1.842483	1.661591
300059	100	0.075344	0.029511	39.16753	0.07435	0.022717	30.55376	1.299068	1.281922
300061	128	0.066224	0.027683	41.80274	0.080788	0.022009	27.24291	1.257827	1.534445
300062	137	0.221002	0.059082	26.73349	0.109712	0.029683	27.05557	1.9904	0.988095
300063	59	0.143299	0.065068	45.40751	0.109886	0.030506	27.76131	2.132978	1.63564
300064	42	0.106121	0.068854	64.88278	0.103436	0.030559	29.54353	2.253179	2.196175
300071	89	0.091101	0.042185	46.30615	0.096548	0.027244	28.21784	1.548432	1.641024
300072	43	0.148178	0.078723	53.12722	0.122072	0.03248	26.60736	2.42372	1.996712
300073	94	0.012347	0.006693	54.20794	0.015075	0.006603	43.8042	1.013635	1.237505
300081	161	0.125046	0.03698	29.57281	0.113701	0.026416	23.23277	1.399909	1.272892
300082	97	0.065402	0.032926	50.34365	0.073602	0.024008	32.6181	1.371475	1.543427
300083	85	0.098959	0.052716	53.27014	0.090554	0.029479	32.55391	1.788241	1.636367
300085	54	0.113969	0.067066	58.84609	0.142093	0.037306	26.25462	1.79773	2.241361
300091	273	0.051495	0.017863	34.68775	0.056801	0.01624	28.59017	1.099935	1.213275
300092	63	0.028079	0.028069	99.9641	0.02654	0.023287	87.74437	1.205343	1.139265
300093	27	0.046932	0.02859	60.91936	0.059662	0.022388	37.52458	1.277047	1.623452
300094	162	0.172964	0.043172	24.96026	0.111249	0.027032	24.29846	1.597089	1.027237
300101	1729	0.101941	0.009813	9.626502	0.099196	0.00955	9.628451	1.02759	0.999798
300102	294	0.068738	0.017647	25.67207	0.07514	0.016136	21.47492	1.093596	1.195444
300103	102	0.095405	0.03752	39.32729	0.110881	0.026819	24.18731	1.399005	1.625947
300104	271	0.067866	0.017605	25.94065	0.07823	0.016062	20.53227	1.096034	1.263409

Table A 21 EBLUP F-H estimates of the headcount ratio (F-H HR), the poverty gap (F-H_PG) and the unemployment rate (F-H_UR) in Greece for the years 2009 and 2013

	Domain	F-H_HR 2013	F-H_HR 2009	F-H_PG 2013	F-H_PG 2009	F-H_UR 2013	F-H_UR 2009
300001	Etolia and Akarnania	35.95875	25.22981	13.38452	5.979207	34.92489	9.153104
300003	Viotia	14.56123	16.70904	1.795033	6.113366	26.15544	6.163159
300004	Evia	25.79379	35.3486	10.22238	12.36241	33.03745	12.2551
300005	Evrytania	18.62276	34.80937	4.557508	10.35043	41.7948	2.012521
300006	Fthiotida	20.29907	21.2686	8.094234	4.720792	32.66494	5.350046
300007	Fokida	6.73383	34.79049	0.757606	8.858122	36.96903	9.042295
300011	Argolida	26.68609	25.10625	9.602847	7.315416	22.66376	3.744571
300012	Arkadia	19.5759	27.91836	5.968121	10.39286	26.03034	0.812001
300013	Achaia	26.4433	25.65865	11.38108	7.283702	35.1014	12.97829
300014	Ilia	21.92731	35.00504	5.958873	11.20806	33.15012	11.13676
300015	Korinthia	24.52514	16.75132	9.7334	6.834966	25.01094	8.133205
300016	Lakonia	19.65326	27.21281	3.025774	8.045864	23.18093	2.815604
300017	Messinia	23.85348	33.27922	8.981779	10.48452	26.16163	4.240069
300021	Zakynthos	28.74388	28.90837	10.86205	8.840169	19.29766	5.192268
300022	Kerkyra	21.38667	12.22942	10.06164	3.778925	23.31381	3.011519
300023	Kefallinia	26.34653	-	10.66847	-	25.19895	8.043718
300024	Lefkada	20.70181	29.23546	2.587716	8.375147	-	-
300031	Arta	24.70777	33.87864	10.79321	8.919637	25.63777	-
300032	Thesprotia	26.98367	29.80928	10.26871	6.103831	14.77011	7.660198
300033	Loannina	23.59206	22.8537	7.370506	3.673499	27.06319	10.48135
300034	Preveza	22.42806	34.37026	9.889194	9.996314	23.30678	10.31192
300041	Karditsa	22.63469	18.98016	9.397696	6.966253	27.72652	7.675719
300042	Larissa	20.83373	20.13857	9.739754	7.906749	26.56953	10.96501
300043	Magnissia	22.94005	15.19497	7.513632	4.200721	31.30867	11.16581
300044	Trikala	21.81313	25.70542	6.830028	8.956544	24.99315	7.528543
300051	Grevena	20.5227	36.6624	4.765667	10.70133	28.55911	4.692405
300052	Drama	26.84941	30.80399	11.64553	10.94761	33.01152	7.966489
300053	Imathia	39.01108	25.06384	17.68134	7.305563	27.86338	10.01929
300054	Thessaloniki	22.17725	18.06205	8.191683	5.066739	33.6468	10.38588
300055	Kavala	20.94791	27.35603	10.00852	6.794159	27.86691	9.969625
300056	Kastoria	23.67259	28.23744	7.694676	10.6501	21.29123	8.249321
300057	Kilkis	21.81619	39.72588	7.898155	13.37886	32.59149	11.03175
300058	Kozani	24.28858	16.13442	7.454425	5.535478	31.63315	12.83239
300059	Pella	31.88734	26.90155	11.42871	8.514249	25.87058	7.43499
300061	Pieria	30.11416	31.32425	11.99665	7.644187	30.10576	8.078754
300062	Serres	23.90497	43.11878	8.212291	13.40169	27.2276	10.97123
300063	Florina	28.27253	33.47161	10.78806	9.799782	27.32904	10.98864
300064	Chalkidiki and Aghion Oros	26.79967	24.96848	6.749395	5.078951	26.70052	10.34365
300071	Evros	21.38057	26.62294	6.526621	7.142049	20.88475	9.654811
300072	Xanthi	38.16437	36.03898	16.79641	8.923152	37.08024	12.20724
300073	Rodopi	32.80927	31.64562	10.38381	9.007799	21.64018	1.507469
300081	Dodekanissos	26.60949	21.66304	7.270967	5.045929	21.27769	11.37006
300082	Kyklades	21.14544	11.87839	6.570589	2.819023	21.81569	7.360192
300083	Lesvos	27.54061	37.95107	10.18021	8.420159	25.80088	9.055435
300084	Samos	17.76952	29.56326	4.097465	7.286829	23.73888	0
300085	Chios	19.98697	21.60007	6.631549	4.746725	26.8133	14.20932
300091	Iraklio	21.34732	13.81004	7.632646	4.185168	22.25287	5.680144
300092	Lassithi	26.72763	10.41253	10.66321	4.701734	13.3956	2.653994
300093	Rethymno	30.01547	4.875683	12.51162	0.622484	25.24985	5.966195
300094	Chania	21.26725	19.11487	5.91956	5.377213	25.61351	11.1249
300101	Prefecture of Athens	19.9659	13.26621	7.254517	3.181138	28.94839	9.919641
300102	Prefecture of East Attiki	22.35602	10.98736	7.457251	4.301431	27.01808	7.513997
300103	Prefecture of West Attiki	32.70915	22.07936	12.47923	7.789471	36.36248	11.08814
300104	Prefecture of Pireas	14.58197	15.5255	5.928262	4.825717	38.47751	7.823022