The University of Manchester

MANCHESTER 1824

Cathie Marsh Centre for Census and Survey Research

M-Quantile Models with Application to Poverty Mapping

CCSR Working Paper 2007-05
Nikos Tzavidis, Nicola Salvati, Monia Pratesi, Ray Chambers
nikos.tzavidis@manchester.ac.uk

Over the last decade there has been growing demand for estimates of population characteristics at small area level. Unfortunately, cost constraints in the design of sample surveys lead to small sample sizes within these areas and as a result direct estimation, using only the survey data, is inappropriate since it yields estimates with unacceptable levels of precision. Small area models are designed to tackle the small sample size problem. The most popular class of models for small area estimation is random effects models that include random area effects to account for between area variations. However, such models also depend on strong distributional assumptions, require a formal specification of the random part of the model and do not easily allow for outlier robust inference. An alternative approach to small area estimation that is based on the use of M-quantile models was recently proposed by Chambers and Tzavidis (2006) and Tzavidis and Chambers (2007). Unlike traditional random effects models, M-quantile models do not depend on strong distributional assumption and automatically provide outlier robust inference.

www.ccsr.ac.uk

M-quantile Models with Application to Poverty Mapping

Nikos Tzavidis
Centre for Census and Survey Research
University of Manchester
Manchester M13 9PL, UK
Nikos.Tzavidis@manchester.ac.uk

Nicola Salvati and Monica Pratesi
Dipartimento di Statistica e Matematica Applicata all'Economia
Università di Pisa
Via Ridolfi 10, Pisa 56124, Italy
salvati@ec.unipi.it
m.pratesi@ec.unipi.it

Ray Chambers
Centre for Statistical and Survey Methodology
School of Mathematics and Applied Statistics
University of Wollongong
Wollongong, NSW 2522, Australia
ray@uow.edu.au

Abstract

Over the last decade there has been growing demand for estimates of population characteristics at small area level. Unfortunately, cost constraints in the design of sample surveys lead to small sample sizes within these areas and as a result direct estimation, using only the survey data, is inappropriate since it yields estimates with unacceptable levels of precision. Small area models are designed to tackle the small sample size problem. The most popular class of models for small area estimation is random effects models that include random area effects to account for between area variations. However, such models also depend on strong distributional assumptions, require a formal specification of the random part of the model and do not easily allow for outlier robust inference. An alternative approach to small area estimation that is based on the use of M-quantile models was recently proposed by Chambers and Tzavidis (2006) and Tzavidis and Chambers (2007). Unlike traditional random effects models, M-quantile models do not depend on strong distributional assumption and automatically provide outlier robust inference. In this paper we illustrate for the first time how M-quantile models can be practically employed for deriving small area estimates of poverty and inequality. The methodology we propose improves the traditional poverty mapping methods in the following ways: (a) it enables the estimation of the distribution function of the study variable within the small area of interest both under an M-quantile and a random effects model, (b) it provides analytical, instead of empirical, estimation of the mean squared error of the M-quantile small area mean estimates and (c) it employs a robust to outliers estimation method.

The methodology is applied to data from the 2002 Living Standards Measurement Survey (LSMS) in Albania for estimating (a) district level estimates of the incidence of poverty in Albania, (b) district level inequality measures and (c) the distribution function of household per-capita consumption expenditure in each district. Small area estimates of poverty and inequality show that the poorest Albanian districts are in the mountainous regions (north and north east) with the wealthiest districts, which are also linked with high levels of inequality, in the coastal (south west) and southern part of country. We discuss the practical advantages of our methodology and note the consistency of our results with results from previous studies. We further demonstrate the usefulness of the M-quantile estimation framework through design-based simulations based on two realistic survey data sets containing small area information and show that the M-quantile approach may be preferable when the aim is to estimate the small area distribution function.

Keywords: Distribution function; Quantile regression; Inequality measure; Poverty assessment; Robust inference

1. Introduction

Over the last decade there has been growing demand for producing estimates of population characteristics at disaggregated geographical levels, often referred to as small areas or small domains. Two popular examples are the estimation of poverty incidence and the estimation of average income at small area level (Bigman *et. al.* 2000). Unfortunately, cost constraints in the design of sample surveys lead to small sample sizes within small areas. As a result direct estimation using only the survey data is inappropriate as it yields estimates with unacceptable levels of precision. In such cases small area estimation can be performed via models that borrow strength by using all the available data and not only the area specific data. The most popular class of models for small area estimation is random effects models that include random area effects to account for between area variation beyond that explained by auxiliary variables (Fay and Herriot, 1979; Battese *et. al.,* 1988; Rao, 2003). However, these models depend on strong distributional assumptions, require a formal specification of the random part of the model and do not easily allow for outlier robust inference. A new approach to small area estimation is based on M-quantile models (Chambers and Tzavidis, 2006; Tzavidis and Chambers, 2007). Unlike random effects small area models, M-quantile small area models do not depend on strong distributional assumptions and outlier robust inference is automatically performed when these models are fitted.

In this paper we illustrate, for the first time, how M-quantile models can be practically employed for deriving small area estimates of poverty and inequality. The methodology we propose improves traditional poverty mapping methods in the following ways: (a) it enables the estimation of the distribution function of the study variable within the small area of interest, (b) it allows analytical, instead of empirical, estimation of the mean squared error of the small area mean estimates and (c) it employs a robust to outliers estimation method. The paper is organized as follows. In section 2 we review linear mixed effects models and M-quantile models for small area estimation and define a unified framework for small area estimation under which target parameters at the small area level are defined as functionals of the small area distribution function. We also review mean squared error estimation for M-quantile estimators of the small area mean. In section 3 we demonstrate the usefulness of our methodological framework through simulations based on two realistic survey data sets containing small area information and show that the M-quantile approach may be preferable when the aim is to estimate the small area distribution function. In section 4 the methodology is applied to data from the 2002 Living Standards Measurement Survey (LSMS) in Albania for estimating (a) district level estimates of the incidence of poverty in Albania, (b) district level inequality measures and (c) the distribution function of household per-capita consumption expenditure in each district. We discuss the practical advantages of our methodology and notice the consistency of our results with results from previous studies. Finally in section 5 we summarize our findings and provide directions for further developments.

2. M-quantile Models for Small Area Estimation

In what follows we assume that a vector of p auxiliary variables x is known for each population unit i in small area j and that information for the variable of interest y is available from a sample which we denote by s that includes units from all the small areas of interest. We denote the population (sample) size in area j by N_j (n_j) and use s_j (r_j) to denote the sampled (non-sampled) population units this area. The target is to use these data to estimate various area specific quantities, including (but not only) the small area mean m_j of y. Linear mixed models are widely used for this purpose. In the general case a linear mixed model for the value of y in area j has the following form

$$y_i = x_i^T \beta + z_i^T \gamma_j + \varepsilon_i,\ i = 1, \ldots, n_j, j = 1, \ldots, d, \tag{1}$$

where γ_j denotes a vector of area level random effects and z_i denotes a vector of auxiliary variables whose values are known for all units in the population. The role of the random effects in the model is to characterize differences in the conditional distribution of y given x between the small areas. Small area estimation under this model is typically based on substituting efficient (e.g. ML or REML) estimates for unknown parameters in (1) under the assumption that all random effects are normally distributed. For example, m_j is estimated via its empirical best linear unbiased predictor (EBLUP)

$$\hat{m}_j = N_j^{-1}\left[\sum_{i \in s_j} y_i + \sum_{i \in r_j}\left(x_i^T \hat{\beta} + z_i^T \hat{\gamma}_j\right)\right] \tag{2}$$

where a ‘hat’ denotes an estimated quantity (see Rao 2003).

An alternative approach to small area estimation is based on the use of M-quantile models (Breckling and Chambers, 1988). M-quantile regression provides a ‘quantile-like’ generalization of regression based on influence functions. A linear M-quantile regression model is one where the q^{th} M-quantile $Q_q(x;\psi)$ of the conditional distribution of y given x satisfies

$$Q_q(x;\psi) = x^T \beta_\psi(q) \tag{3}$$

with ψ denoting the influence function associated with the M-quantile. That is, we allow a different set of regression parameters for each value of q. For specified q and continuous ψ, an estimate $\hat{\beta}_\psi(q)$ of $\beta_\psi(q)$ can be obtained via an iterative weighted least squares algorithm.

Chambers and Tzavidis (2006) extended the use of M-quantile models to small area estimation. Following their development, the conditional variability across the population of interest is characterised by the M-quantile coefficients of the population units. For unit i with values y_i and x_i, this coefficient is the value q_i such that $Q_{q_i}(x_i;\psi) = y_i$. If a hierarchical structure does explain part of the variability in the population data, then we expect units within groups (areas) defined by this hierarchy to have similar M-quantile coefficients. Under this approach, we avoid specifying random effects. Instead, intra-area variation is captured via the area specific M-quantile coefficients, which in this paper are calculated as the average value $\hat{\theta}_j$ of the estimated M-quantile coefficients of the sample units in area j. Note that alternative definitions of $\hat{\theta}_j$ are possible, e.g. sample-weighted averages, but they will not be considered in this paper.

Under this approach, the area level coefficient $\hat{\theta}_j$ characterises the behaviour of y in area j. Thus, Chambers and Tzavidis (2006) suggest that m_j be estimated by

$$\hat{m}_j = N_j^{-1}\left[\sum_{i\in s_j} y_i + \sum_{i\in r_j} x_i^T \hat{\beta}_\psi(\hat{\theta}_j)\right]. \qquad (4)$$

Similarly, they suggest that the p^{th} quantile m_{pj} of the distribution of y in area j be estimated by the solution to

$$\int_{-\infty}^{m_{pj}} d\hat{F}_j(t) = p \qquad (5)$$

where $\hat{F}_j(t)$ denotes the 'plug-in' estimate of the area j distribution function

$$\hat{F}_j(t) = N_j^{-1}\left[\sum_{i\in s_j} I(y_i \le t) + \sum_{i\in r_j} I\left\{x_i^T \hat{\beta}_\psi(\hat{\theta}_j) \le t\right\}\right]. \qquad (6)$$

Tzavidis and Chambers (2007) note that both (4) and (5) can be biased and propose alternative estimators based on replacing (6) by the distribution function estimator proposed by Chambers and Dunstan (1986)

$$\hat{F}_{CD,j}(t) = N_j^{-1}\left[\sum_{i\in s_j} I(y_i \le t) + n^{-1}\sum_{i\in s_j}\sum_{k\in r_j} I\left\{\hat{y}_k + \left(y_i - \hat{y}_i\right) \le t\right\}\right] \qquad (7)$$

where $\hat{y}_i = x_i^T \hat{\beta}_\psi(\hat{\theta}_j)$ when population unit i is from area j. This leads to an alternative to (4) of the form

$$\hat{m}_j = \int_{-\infty}^{\infty} t d\hat{F}_{CD,j}(t) = N_j^{-1}\left[\sum_{i\in s_j} y_i + \sum_{i\in r_j} x_i^T \hat{\beta}_\psi(\hat{\theta}_j) + \frac{N_j - n_j}{n_j}\sum_{i\in s_j}\left(y_i - \hat{y}_i\right)\right]. \qquad (8)$$

The p^{th} quantile m_{pj} is still computed by solving (5), but now with the plug-in estimator (6) replaced by the Chambers-Dunstan (CD) estimator (7). They also note that by substituting $\hat{y}_i = x_i^T \hat{\beta} + z_i^T \hat{\gamma}_j$ in (7) one can define mixed model versions of the CD-based mean estimator (8) as well as corresponding estimators of the within area quantiles of y.

Mean squared error (MSE) estimation of M-quantile based small area mean estimators relies on the approach described in Chandra and Chambers (2005) and Chambers and Tzavidis (2006). Since the estimates $\hat{\beta}_\psi(q)$ of the M-quantile regression coefficients can be expressed as linear combinations of the sample y values, it follows that, for fixed $\hat{\theta}_j$, the estimator of the area j M-quantile regression coefficient is also linear in these values. Consequently, both the plug-in version (4) and the CD-based version (8) of the M-quantile estimator of m_j can be written as linear combinations of these sample values and a first order approximation to their mean squared errors developed using the arguments in Royall and Cumberland (1978). Let $\{w_i; i \in s\}$ denote the set of weights that 'define' either (4) or (8). This approach then leads to a MSE estimator of the form

$$\hat{M}_j = \hat{V}(\hat{m}_j) + \left\{ \hat{B}(\hat{m}_j) \right\}^2 \tag{9}$$

where

$$\hat{V}(\hat{m}_j) = N_j^{-2} \sum_g \sum_{i \in s_g} \left[\left\{ (w_i - 1)^2 + \frac{N_j - n_j}{n_j - 1} \right\} I(g = j) + w_i^2 I(g \neq j) \right] \left\{ y_i - x_i^T \hat{\beta}_\psi(\hat{\theta}_g) \right\}^2$$

and

$$\hat{B}(\hat{m}_j) = N_j^{-1} \left[\sum_g \sum_{i \in s_g} w_i x_i^T \hat{\beta}_\psi(\hat{\theta}_g) - \sum_{i \in j} x_i^T \hat{\beta}_\psi(\hat{\theta}_j) \right].$$

Note that if $\hat{m}_j$ is defined to be the CD-based estimator (8), then $\hat{B}(\hat{m}_j)$ is zero.

In this paper we restrict ourselves to providing mean squared error estimates for means using (9). Extension of this approach to estimation of the mean squared error of area level quantile estimates is technically feasible via a linearisation argument, but remains untested. Numerically intensive methods for mean squared error estimation (e.g. the bootstrap) are also possible.

3. Why Use M-Quantile Models for Small Area Estimation?

Chambers and Tzavidis (2006) discuss a number of practical reasons for preferring M-quantile models over mixed models for small area estimation. In large part, these advantages relate to the flexibility and simplicity of M-quantile modelling (e.g. there is no need to specify a model for the random effects) as well as the inbuilt robustness of the M-quantile approach to the distributional assumptions that are inherent in mixed models. Some appreciation for this robustness can be obtained from the results of two simulation studies presented in Tzavidis and Chambers (2007), and which we now summarise.

The first study is based on a population of 81982 Australian broadacre farms, created by bootstrapping a sample of 1652 such farms. These farms are spread across 29 Australian broadacre farming regions, and the aim is to estimate the average annual farm costs in each region. The simulation was carried out by repeatedly sampling 500 times from this population using a stratified random sampling design with strata defined by the regions and with stratum sizes fixed to be the same as those in the original sample. The second study again used a bootstrap population, this time made up of 724782 households and based on a sample of 3591 households spread across the 36 districts of Albania that provided data for the 2002 Living Standards Measurement Survey. The variable of interest here was per-capita consumption expenditure. The simulation design used (stratified random sampling with districts as strata) was similar to that in the first study, but in this case 200 simulations were carried out. A noteworthy feature of the Australian population is its pronounced heteroskedasticity as well as the presence of a number of very large outliers. In contrast the Albanian population is much better behaved, with no obvious outliers. Tables 1 and 2 show the biases and root mean squared errors recorded in these two simulation studies for the EBLUP (2) based on a random intercepts version of (1), the plug-in M-quantile estimator (4) and the CD-based M-quantile estimator (8). Focusing on Table 1, we see that when the target of inference is the small area mean the plug-in M-quantile estimator (4) is severely biased. The bias reduces significantly once we use the CD-based estimator (8). Note that this estimator is also more efficient than the EBLUP in this situation. When the target of inference is the small area median the plug-in M-quantile estimator (4) is the best performer. However, the CD-based estimator (8) dominates it as well as the EBLUP, which in this case is calculated using (5) and (6) but with fitted values derived from (1), at all other quantiles. Further results, not reported in this paper, show that the CD-based M-quantile estimator and the CD-based EBLUP estimator have similar performance when the target of inference is the small area median. Figure 1 shows the distribution of actual coverage rates of nominal 95 per cent confidence intervals for regional means derived using the CD-based M-quantile estimator (8) and the mean squared error estimator (9). In general these confidence intervals display coverage rates close to nominal levels, with significant under-coverage only in one region containing a very large outlier.

Table 2 displays similar simulation results for estimation of district averages under the M-quantile and EBLUP estimators. Here we see again the bias in the plug-in M-quantile estimator, with the CD-based M-quantile and the EBLUP estimators performing similarly, both showing low bias and essentially equivalent mean squared errors. In Figure 1 we see the distribution of actual coverage rates of nominal 95 per cent confidence intervals for district averages derived using the CD-based M-quantile estimator (8) and the mean squared error estimator (9). Coverage rates are above 80 per cent for all 36 districts, with only four districts recording coverage rates less than 90 per cent.

Additional simulation results can be found in Tzavidis and Chambers (2007). These show that the CD-based M-quantile estimator (8), with its mean squared error estimator (9), performs consistently well.

4. Poverty Mapping for Albania Using M-quantile Models

Poverty maps are important tools that provide information on the spatial distribution of poverty and are often used to assist the implementation of poverty alleviation programmes. Poverty can be defined both in terms of income deprivation and inadequacies in a number of non-income measures of welfare such as education, health and access to basic services and infrastructures. In this study we refer to the poverty as income deprivation. Since the economy of Albania is largely rural and income is not accurately measured, income poverty in Albania is estimated on the basis of a consumption-based measure (World Bank, 2003). A household is defined as poor if its per-capita consumption expenditure falls below a minimum level (poverty line) necessary to meet the basic food and non-food needs. Deriving poverty estimates also depends on how one sets the poverty line. The choices in setting the poverty line can impact upon the estimates and hence have an effect on policy making. For this reason it is important to obtain a picture of the distribution function of per-capita consumption expenditure within each small area (district).

4.1 The Data

We employ data from the Living Standards Measurement Study (LSMS) that was conducted in 2002 in Albania. The LSMS was established by the World Bank in 1980 to explore ways of improving the type and quality of household data collected by statistical offices in developing countries. Its goal is to foster the increased use of household data as a basis for policy decision making. This survey provides valuable information on a variety of issues related to living conditions of the people in Albania, including details on income and non-income dimensions of poverty in the country, and forms the basis of poverty assessment in this country. There are twelve prefectures in Albania with a prefecture consisting of several districts. There are thirty six districts in total (Figure 2). An attempt to obtain direct estimates of household per-capita consumption expenditure at district level reveals the lack of precision (increased variance) of the direct estimates particularly for districts with small sample sizes.

4.2 Data Analysis and Results

In order to produce measures that describe the spatial distribution of poverty and inequality at disaggregated geographical level, for example at district level, the World Bank employs the approach of Elbers, Lanjouw and Lanjouw (2003), hereafter referred to as the ELL method. The idea underpinning the ELL method is to estimate a linear regression model using the logarithmic transformation of the household per-capita consumption expenditure with local variance components. At the first stage this regression model is estimated using the survey data. At the second stage the estimated model parameters are combined with census data for simulating the per-capita consumption expenditure of each household in the population. These household level predictions are then used to form estimates of the incidence of poverty. In this paper we suggest an alternative approach for estimating the incidence of poverty at disaggregated geographical levels that is based on the use of small area models. Under this approach the parameters of a small area model, such as a random effects or an M-quantile model, are estimated using the survey data. The estimated model parameters are then combined with census level covariate information to form predictions of per-capita consumption expenditure for each household in the population. These household level predictions are compared to the poverty line for deriving local estimates of the incidence of poverty. Although poverty mapping can be performed using either M-quantile or random effects models, in this application we exclusively focus on the use of M-quantile models.

We employ an M-quantile small area model for estimating (a) the incidence of poverty and inequality in Albanian districts and (b) the distribution function of household per-capita consumption expenditure in each district. The methodology we propose improves the traditional poverty mapping methods in the following ways: (a) it enables the estimation of the distribution function of the study variable within the small area of interest, (b) it provides analytical, instead of empirical, estimation of the mean squared error of the estimates of the small area mean and (c) it employs a robust to outliers estimation method. Although we do not show results here, the approach we take also allows estimation of the small area distribution functions of the study variable under a model that characterizes small area differences via random effects.

The selection of covariates to fit the M-quantile model relies on prior studies of poverty assessment in Albania (Betti, 2003). While selecting covariates it is important to consider the various non-income dimensions of poverty and deprivation that can potentially dominate the pure income dimension. We have selected the following household level covariates: the household size, which is a strong indicator of poverty, the presence of facilities in the dwelling (TV, parabolic dish antenna, refrigerator, air conditioning, personal computer), ownership of dwelling, ownership of land and ownership of car.

Poverty indicators: To measure poverty, a set of poverty indices has been used. These include the headcount ratio (HCR) and two indices that belong to the family of measures proposed by Foster, Green and Thorbecke (1984), hereafter referred to as FGT. These are the poverty gap index, denoted FGT(1), and the poverty severity index, denoted FGT(2). The headcount ratio measures the proportion of individuals or households in the population that are poor as this is defined by the poverty line. The poverty line used to obtain the poverty estimates is set equal to 4,891 Leks per month (World Bank, 2003). FGT(1) shows the extent to which individuals fall below the poverty line and expresses this as a percentage of the poverty line. FGT(2) is a weighted sum of the poverty gaps where the weights are the proportionate poverty gaps themselves. This measure attempts to capture the inequality among the poor. District level measures of the incidence of poverty are reported in Table 3. In addition, Figures 3(a) and 3(b) map the estimates of the proportion of poor households and individuals in each of the Albanian districts respectively.

Inequality measures: To estimate inequality we employ the following measures: (a) the Gini coefficient, (b) two measures that belong to the Generalized Entropy (GE) class of decomposable inequality measures namely, GE(α) with α set to 0 and to 1, and (c) the interquartile range (IQR) and the interdecile range (IDR). The Gini coefficient ranges between 0 and 1 with 0 indicating perfect equality and 1 perfect inequality. The values of the GE measures vary between 0 and ∞ with zero representing an equal distribution and higher values representing higher values of inequality. The interquartile range (IQR) is defined as the difference between the 75th and 25th estimated percentiles of the within district distribution function of household per-capita consumption expenditure. Similarly, the interdecile (IDR) range is defined as the difference between the 90th and 10th estimated deciles of the within district distribution function of household per-capita consumption expenditure. Our methodology enables computation of the last two inequality measures because it allows for the estimation of the district level distribution function of the study variable. District level inequality measures are reported in Table 4.

Small area estimates: Estimates of average household per-capita consumption expenditure in each district are derived using the CD-based M-quantile estimator (8) with associated standard errors derived using the mean squared estimator (9). In Table 5 we report 95% confidence intervals for estimates of the average of household per-capita consumption expenditure by district. Estimates of the percentiles of the distribution function of household per-capita consumption expenditure within each district are obtained by numerically integrating (7). These estimates are mapped in Figures 4, 5 and 6 with the actual numbers reported in Table 6.

An examination of the tables of results and of the corresponding maps reveals the districts with higher levels of poverty. The districts of Bulqize (poverty head count ratio of 68%), Puke (56%), Kurbin (49%) and Peqin (44%) are the poorest ones. The district of

Sarande has the lowest percentage of individuals below the poverty line (5%). This is followed by Permet (8%), Gjirokaster (10%) and Vlore (11%). This ranking is further confirmed by the FGT(1) and FGT(2) measures. According to the three poverty indices we considered, the incidence of poverty is higher for districts in the mountainous region of Albania (north and north-east of the country). In contrast, the incidence of poverty is lower for the south western (coastal area) and southern districts of Albania.

According to the inequality measures, the district of Peqin has the highest inequality with the Gini coefficient, GE(0) and GE(1) equal to 38.30%, 25.55% and 26.66%, respectively. For other districts with high poverty rates, Bulqize, Kurbin and Puke, the Gini coefficients are 28.50%, 23.53% and 25.37% respectively. On the other hand the district of Sarande has the lowest value of Gini coefficient (19.46%). For other districts with low poverty rates, Permet, Gjirokaster and Vlore, the Gini coefficients are 23.78%, 24.49 and 25.43% respectively. The results indicate that the levels of inequality for wealthier districts can be equal or higher to the levels of inequality in poorer districts. This is also confirmed by examining the GE(0) and GE(1) measures. Further insight about poverty and inequality in Albania is offered by studying the small area estimates of the household per-capita consumption expenditure. Although the district of Gjirokaster has the highest average per-capita consumption expenditure with 12,219 Leks per month, it does not have the lowest percentage of households and individuals below the poverty line (6% and 10% respectively). This situation calls for a deeper study of the within district distribution function. Our method offers an easy way of doing this. Examining the values of the interquartile and the interdecile ranges we note that for the district of Gjirokaster these values are among the highest. This indicates that one of the wealthiest districts of Albania also has high levels of inequality. The same is also true for the coastal district of Vlore and the southern district of Permet, i.e. relatively low levels of headcount poverty, high average household per-capita consumption expenditure and also high levels of inequality. Finally, the capital district (Tirane) has a moderate poverty rate (21%) and high levels of inequality. The Gini coefficient is equal to 30.34% and the interdecile range is the highest one and equal to 13828.99.

At this point we need to clarify that our results are not directly comparable with results obtained from the ELL method. Firstly this is because we didn't have access to the complete database employed for producing poverty estimates with the ELL method and secondly because the ELL method is applied on the logarithmic scale whereas our approach is applied on the raw scale. Estimating the model on the logarithmic scale requires back transforming to the original scale. This back transformation, however, may result in bias (Dorfman and Chambers, 2003). Nevertheless, our analysis leads to results that are consistent with view of the World Bank and with results obtained by applying the ELL method. M-quantile small area estimates of poverty and inequality confirm that the poorest Albanian districts are in the mountainous regions with the wealthiest districts, which are also linked to high levels of inequality, in the coastal and southern part of country.

5. Concluding Remarks and Further Developments

In this paper we demonstrate how M-quantile models can be employed successfully for estimating the incidence of poverty and inequality at disaggregated geographical levels and for estimating small area distribution functions of the household per-capita consumption expenditure. In doing so, we employ the methodology of Chambers and

Tzavidis (2006) and Tzavidis and Chambers (2007). The application of M-quantile models for poverty mapping in Albania illustrates the potential of employing small area methods in poverty evaluation studies. There are a number of research questions that remain to be tackled. Firstly, a comparison between the ELL approach and the M-quantile approach to poverty mapping is needed. Furthermore, mean squared error estimation for quantiles of small area distribution functions remains to be developed. Finally, information on the geographical location of the sample units, sometimes available in survey data, is often not taken into account. In many practical situations such as in environmental, epidemiological and economic applications the spatial dimension of the data must be explicitly modelled. We are currently studying the use of random effects models with spatially correlated random effects and developing spatial locally robust (M-quantile) models for small area estimation and poverty mapping.

ACNOWLEDGEMENT

The authors wish to thank the editor, the associate editor and the referees for helpful comments that greatly improved the presentation. The first author also wishes to thank the Department of Statistics and Mathematics Applied to Economics, University of Pisa for support in conducting the research described in this paper.

References

Battese GE, Harter RM and Fuller WA (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association,* **83**, No. 401: 28-36.

Betti G (2003). Poverty and Inequality Mapping in Albania: Final Report. *World Bank and INSTAT (mimeo)*, Washington DC and Tirana.

Bigman D, Dercon S, Guillaume D and Lambotte M (2000). Community Targeting for Poverty Reduction in Burkina Faso. *The World Bank Economic Review*, **14** (1): 167-93.

Breckling J and Chambers R (1988). M-quantiles. *Biometrika*, **75** (4): 761-71.

Chambers R and Dunstan R (1986). Estimating Distribution Functions from Survey Data. *Biometrika*, **73** (3): 597-604.

Chambers R and Dorfman AH (2003). Transformed Variables in Survey Sampling. S3RI *Methodology Working Papers*, **M03/21,** Southampton Statistical Sciences Research Institute, University of Southampton, UK.

Chambers R and Tzavidis N (2006). M-quantile Models for Small Area Estimation. *Biometrika*, **93** (2): 255-68.

Chandra H and Chambers R (2005). Comparing EBLUP and C-EBLUP for Small Area Estimation. *Statistics in Transition*, **7** (3): 637-48.

Elbers C, Lanjouw J and Lanjouw P (2003). Micro-Level Estimation of Poverty and Inequality. *Econometrica*, **71** (1): 355-64.

Fay RE and Herriot RA (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**, No.366: 269-77.

Foster J, Greer J and Thorbecke E (1984). A Class of Decomposable Poverty Measures. *Econometrica ,* **52** (3): 761-66.

Rao JNK (2003). Small Area Estimation, John Wiley & Sons, New York.

Ravallion M. (2002). Poverty Lines in Theory and Practice. The World Bank, *LSMS Working Paper* **No. 133**, Washington DC

Royall R.M. and Cumberland W.G. (1978). Variance Estimation in Finite Population Sampling. *Journal of the American Statistical Association*, **73** No 362: 351-58.

Tzavidis N and Chambers R (2007). Robust Prediction of Small Area Means and Distributions, *Working Paper* [*available upon request*].

World Bank (2003). Albania Poverty Assessment. *World Bank Report* **No. 26213-AL**, Washington DC.

Table 1: Simulation results for the Australian farms dataset. Entries show averages, over simulations and regions, of Relative Bias (%) and Relative RMSE (RRMSE, %) of estimates of the average as well as percentiles of the regional distribution of annual farm costs under the different methods.

Method	**Target Percentile**					
	10	**25**	**50**	**Mean**	**75**	**90**
Relative Bias						
EBLUP	120.96	77.44	32.01	4.04	4.15	-4.30
Plug-in M-quantile	79.92	35.63	4.47	-16.16	-20.20	-32.33
CD-based M-quantile	-14.10	19.27	20.48	-0.20	7.29	0.77
RRMSE						
EBLUP	158.41	98.53	46.63	19.60	22.53	27.99
Plug-in M-quantile	88.13	44.98	22.53	20.41	26.82	36.74
CD-based M-quantile	39.79	31.10	29.26	18.23	19.50	20.95

Table 2. Simulation results for the Albanian households dataset. Entries show averages, over simulations and districts, of Relative Bias (%) and Relative RMSE (RRMSE, %) of estimates of district level average household per-capita consumption expenditure under the different methods.

Method	**Relative Bias**
EBLUP	0.61
Plug-in M-quantile	-10.74
CD-based M-quantile	0.07
Method	**RRMSE**
EBLUP	5.46
Plug-in M-quantile	13.30
CD-based M-quantile	5.55

Figure 1. Distribution over small areas of actual coverage rates of nominal 95 per cent confidence intervals for small area means. Intervals in both cases were defined by the CD-based M-quantile estimator (8) plus or minus twice its estimated root mean squared error using (9).

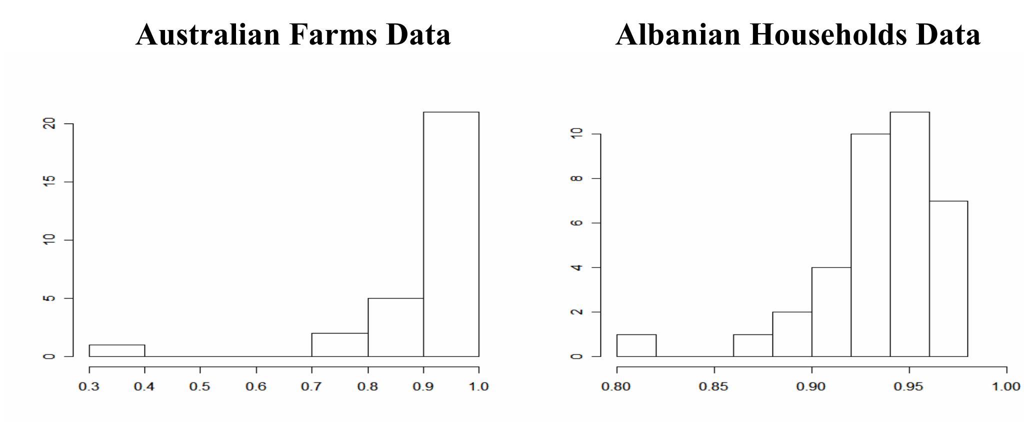

Table 3. Albanian LSMS data: District level estimates of the incidence of poverty (%).

District	Household HCR	Individual HCR	FGT(1)	FGT(2)
BERAT	22	28	8.78	5.62
BULQIZE	57	68	20.83	11.69
DELVINE	12	18	3.82	1.96
DEVOLL	10	13	3.88	2.28
DIBER	16	24	5.23	2.94
DURRES	23	31	8.32	5.55
ELBASAN	18	25	5.43	2.89
FIER	12	16	3.30	1.65
GRAMSH	26	34	9.65	6.12
GJIROKASTER	6	10	1.95	1.16
HAS	13	19	4.95	3.28
KAVAJE	12	17	5.33	3.96
KOLONJE	18	26	3.31	1.03
KORCE	12	18	3.32	1.60
KRUJE	28	36	10.22	5.98
KUCOVE	21	27	5.98	3.24
KUKES	19	29	6.95	4.35
KURBIN	38	49	11.53	5.82
LEZHE	13	18	5.97	4.76
LIBRAZHD	27	35	11.24	7.92
LUSHNJE	14	19	4.89	3.07
MALESI E MADHE	19	27	5.90	3.14
MALLAKASTER	16	21	5.28	2.87
MAT	15	23	5.51	3.73
MIRDITE	19	28	7.63	5.02
PEQIN	37	44	13.15	7.62
PERMET	5	8	1.25	0.60
POGRADEC	21	28	5.94	2.97
PUKE	40	56	14.90	9.08
SARANDE	3	5	0.66	0.31
SKRAPAR	19	24	7.36	4.65
SHKODER	12	19	3.60	2.02
TEPELENE	23	29	9.07	5.20
TIRANE	15	21	5.53	3.58
TROPOJE	23	34	8.20	4.96
VLORE	7	11	2.19	1.25

Table 4. Albanian LSMS data: Inequality indices by district.

District	Gini	GE(0)	GE(1)	IQR	IDR
BERAT	30.48	17.37	14.72	5576.66	11052.36
BULQIZE	28.50	15.01	12.18	3055.53	5933.84
DELVINE	27.63	13.56	11.94	6487.30	13559.44
DEVOLL	24.20	11.53	9.71	5339.45	10971.64
DIBER	24.13	11.05	9.39	4268.83	8520.21
DURRES	28.06	13.98	11.89	4563.85	9409.05
ELBASAN	28.31	14.72	16.00	4275.30	9081.06
FIER	25.64	11.64	11.23	4935.42	10428.14
GRAMSH	29.61	16.24	14.55	4047.68	8522.86
GJIROKASTER	24.49	11.03	10.05	6353.29	11887.23
HAS	24.76	12.05	9.44	5608.75	10868.00
KAVAJE	25.70	12.57	10.14	5342.22	11245.08
KOLONJE	20.04	6.51	6.45	2643.37	6915.67
KORCE	24.03	10.06	9.66	3962.36	9044.36
KRUJE	27.08	13.38	11.18	4235.35	8254.12
KUCOVE	23.49	11.16	8.76	4615.50	7953.00
KUKES	26.18	12.76	10.80	4672.15	9036.35
KURBIN	23.53	10.32	8.56	3033.54	6218.53
LEZHE	28.60	15.68	12.62	6802.02	13297.71
LIBRAZHD	30.21	16.55	13.38	5111.40	9993.05
LUSHNJE	25.81	12.41	10.71	5081.53	10204.59
MALESI E MADHE	27.27	13.47	12.23	4412.24	9364.82
MALLAKASTER	29.00	15.24	13.10	7124.22	12719.81
MAT	23.48	10.69	8.69	4441.34	8283.70
MIRDITE	24.35	12.66	9.21	5258.66	8934.15
PEQIN	38.30	25.55	26.66	4465.55	10232.01
PERMET	23.78	9.67	8.80	6903.64	12932.28
POGRADEC	26.63	12.51	11.22	5084.65	9091.74
PUKE	25.97	13.85	10.24	3510.37	6271.66
SARANDE	19.46	6.49	6.12	5181.41	9882.47
SKRAPAR	22.39	11.95	8.40	4253.17	7665.91
SHKODER	23.69	10.03	9.10	4372.63	8857.85
TEPELENE	28.50	15.52	12.69	5324.03	10671.02
TIRANE	30.34	16.22	14.93	6736.00	13828.99
TROPOJE	27.06	13.48	11.54	4022.78	8552.36
VLORE	25.43	11.53	10.89	5813.68	12246.16

Table 5. Albanian LSMS data: 95% confidence intervals for the mean of household per-capita consumption expenditure by district.

District	Lower Bound	Mean	Upper Bound
BERAT	6422.38	8331.06	10239.73
BULQIZE	3964.69	4778.07	5591.44
DELVINE	7793.95	10165.35	12536.74
DEVOLL	8450.85	10742.24	13033.64
DIBER	7279.82	8156.45	9033.09
DURRES	6419.35	7641.22	8863.10
ELBASAN	7114.69	8301.25	9487.80
FIER	7397.85	9282.87	11167.88
GRAMSH	6433.02	7326.16	8219.30
GJIROKASTER	10328.36	12219.40	14110.43
HAS	8301.39	9785.44	11269.49
KAVAJE	8345.91	9630.30	10914.70
KOLONJE	5189.91	6919.38	8648.86
KORCE	7746.40	8623.06	9499.71
KRUJE	5789.98	6832.48	7874.99
KUCOVE	6560.63	7644.68	8728.72
KUKES	7156.29	8024.30	8892.31
KURBIN	5000.39	5745.14	6489.89
LEZHE	9317.08	10795.43	12273.77
LIBRAZHD	6689.04	7502.92	8316.81
LUSHNJE	7882.03	9058.31	10234.59
MALESI E MADHE	6325.20	8240.21	10155.21
MALLAKASTER	7605.74	9540.27	11474.79
MAT	7187.81	8262.86	9337.90
MIRDITE	6610.47	8061.28	9512.09
PEQIN	4808.15	7423.50	10038.85
PERMET	9027.05	11303.17	13579.29
POGRADEC	6902.81	8021.51	9140.22
PUKE	4161.45	5538.30	6915.15
SARANDE	10073.17	11342.82	12612.48
SKRAPAR	6186.42	7556.59	8926.75
SHKODER	7285.88	8796.42	10306.95
TEPELENE	6588.81	7961.71	9334.60
TIRANE	7302.76	9596.29	11889.81
TROPOJE	6452.37	7526.35	8600.33
VLORE	8667.40	11205.66	13743.91

Table 6. Albanian LSMS data: Estimates of the percentiles of the distribution of per-capita consumption expenditure by district.

District	Percentile				
	10	25	50	75	90
BERAT	3128.77	5135.13	7629.47	10711.79	14181.13
BULQIZE	1966.38	3099.46	4466.35	6154.99	7900.22
DELVINE	4464.13	6392.56	9141.77	12879.86	18023.57
DEVOLL	4986.41	7728.29	10475.63	13067.74	15958.05
DIBER	3969.01	5805.52	8075.34	10074.35	12489.22
DURRES	3253.40	5031.02	7158.64	9594.87	12662.45
ELBASAN	3846.66	5455.79	7414.92	9731.09	12927.72
FIER	4783.15	6214.58	8217.88	11150.00	15211.29
GRAMSH	3128.76	4836.58	6670.37	8884.26	11651.62
GJIROKASTER	5957.92	8506.14	11777.96	14859.43	17845.15
HAS	4461.57	7116.74	9797.34	12725.49	15329.57
KAVAJE	4318.46	6780.67	9229.32	12122.89	15563.54
KOLONJE	4393.22	5240.52	6214.28	7883.89	11308.89
KORCE	4622.66	6100.83	7991.49	10063.19	13667.02
KRUJE	2988.13	4615.16	6459.10	8850.51	11242.25
KUCOVE	3800.78	5257.03	7591.28	9872.53	11753.78
KUKES	3624.51	5475.31	7569.16	10147.46	12660.86
KURBIN	2874.72	4162.85	5593.04	7196.39	9093.25
LEZHE	4164.09	6970.98	10160.73	13773.00	17461.80
LIBRAZHD	2669.66	4663.63	7164.63	9775.03	12662.71
LUSHNJE	4228.67	6182.29	8554.80	11263.82	14433.26
MALESI E MADHE	3779.79	5416.20	7460.97	9828.44	13144.61
MALLAKASTER	4078.69	6018.94	8251.58	13143.16	16798.50
MAT	4025.78	5948.81	8024.38	10390.15	12309.48
MIRDITE	3476.03	5627.70	8116.96	10886.36	12410.18
PEQIN	2527.51	3956.12	5783.93	8421.67	12759.52
PERMET	5729.37	7624.94	10468.87	14528.58	18661.65
POGRADEC	3789.28	5143.47	7323.93	10228.12	12881.02
PUKE	2390.40	3710.76	5450.77	7221.13	8662.06
SARANDE	6816.06	8490.54	10845.07	13671.95	16698.53
SKRAPAR	3409.03	5484.15	7748.40	9737.32	11074.94
SHKODER	4525.86	6231.09	8367.32	10603.72	13383.71
TEPELENE	2951.18	5110.98	7570.79	10435.01	13622.20
TIRANE	3129.01	5144.00	8017.00	11880.00	16958.00
TROPOJE	3292.88	5119.64	7192.06	9142.42	11845.24
VLORE	5504.34	7608.08	10364.06	13421.76	17750.50

Figure 2. Albanian districts.

Figure 3. District level estimates of Head Count Ratio (a) at household and (b) at individual level.

(a) **(b)**

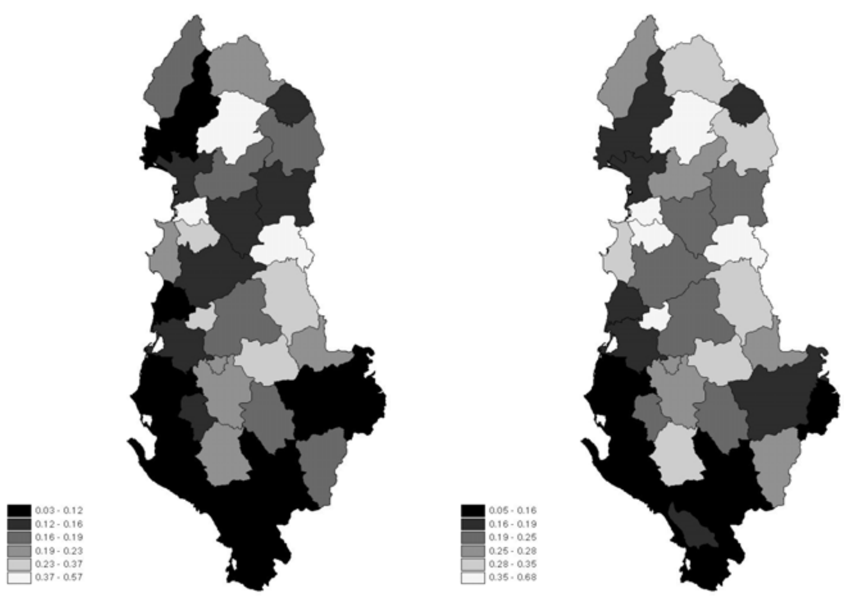

Figure 4. District level estimates of the (a) average per-capita consumption expenditure and (b) median per-capita consumption expenditure.

(a) **(b)**

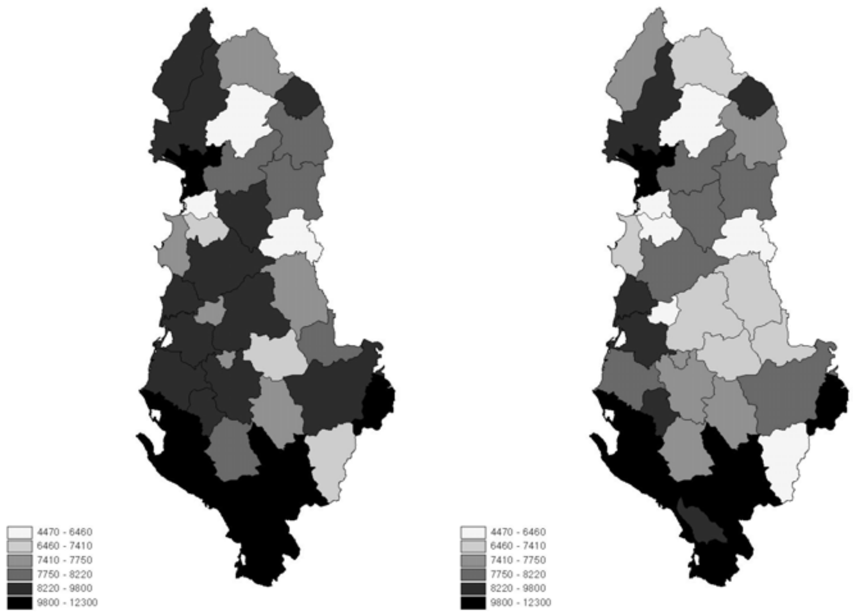

Figure 5. District level estimates of the (a) 10th percentile of per-capita consumption expenditure and (b) 25th percentile of per-capita consumption expenditure.

(a) **(b)**

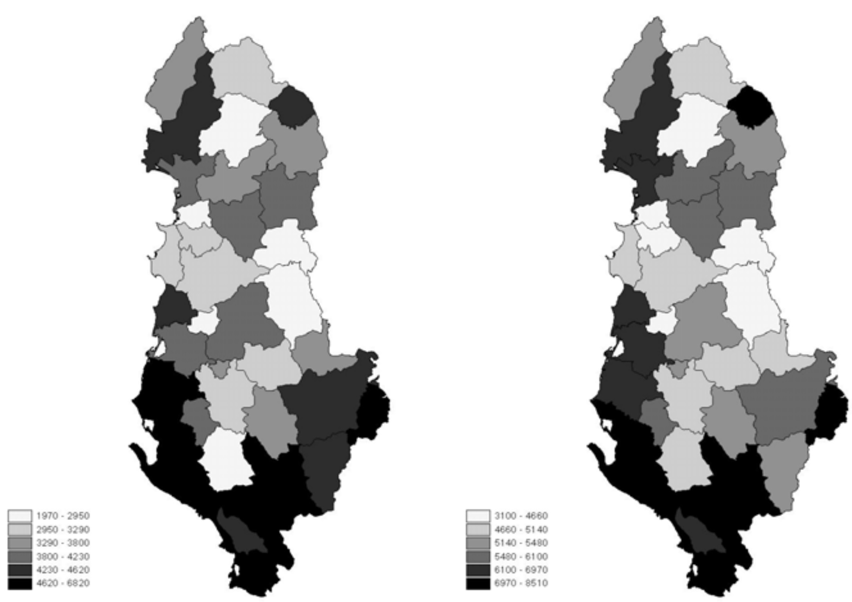

Figure 6. District level estimates of the (a) 75th percentile of per-capita consumption expenditure and (b) 90th percentile of per-capita consumption expenditure.