METHODS BRIEFING 21

M-quantile Models for Small Area Estimation

Ray Chambers* and Nikos Tzavidis**
*Centre for Centre for Statistical and Survey Methodology,
School of Mathematics and Applied Statistics
University of Wollongong, Australia
**Centre for Longitudinal Studies, University of London, UK

March 2003 - June 2006

A series of methods briefings from projects funded by ESRC as part of the Research Methods Programme.

The Programme aims to develop qualitative and quantitative methods within the context of substantive research. It also aims to encourage effective dissemination of good practice.

**Further copies are available from:
CCSR
2nd Floor
Crawford House
Manchester
M13 9PL**

0161 275 4891

www.ccsr.ac.uk/methods/

Background and aims of the research

Sample surveys provide a cost effective way of obtaining estimates for characteristics of interest at both population and subpopulation (domain) level. An estimator of a domain characteristic is called direct if it is based only on data from sample units in the domain. In most practical applications, however, domain sample sizes are not large enough to allow direct estimation; the term 'small areas' is typically used to describe such domains. When direct estimation is not possible, one has to rely upon alternative methods that depend on the availability of population level auxiliary information and are commonly referred to as indirect or model-based methods.

Model-based methods can be classified into two categories, namely methods based on fixed effects models, i.e. models that explain between-area variation in the target variable using only the auxiliary information, and methods based on mixed (random) effects models that include area-specific random effects to account for between-area variation beyond that explained by the auxiliary information. Mixed effects models are widely used in small area estimation (Rao, 2003, Ch.5). However, such models depend on parametric and distributional assumptions as well as requiring specification of the random part of the model. Furthermore, robust inference under these models is not well understood.

In this project we propose a new approach to small area estimation based on modelling quantile-like coefficients of the conditional distribution of the target variable given the covariates. Random effects are avoided. Instead, inter-domain variation is characterised by variation in area-specific values of these quantile-like coefficients. It is easy to fit these new models and they have a number of practical advantages, including easy nonparametric specification and straightforward incorporation of survey weights. Outlier-robust inference is also uncomplicated, as is estimation of other small area characteristics, such as medians and percentiles.

February 2007

The project had several aims and objectives

- To investigate the use of multiquantile models as an alternative to multilevel models for small area estimation
- To apply multiquantile methodology for estimation of averages in small areas
- To build models for the distribution, rather than average values, of the survey data
- To develop appropriate diagnostics for random effects in multilevel models

Methods

Random effects models assume that variability associated with the conditional distribution of *y* given *x* can be at least partially explained by a pre-specified hierarchical structure, such as the small areas of interest. The hierarchical structure of data results in both within group and between group heterogeneity. Random effects models aim to

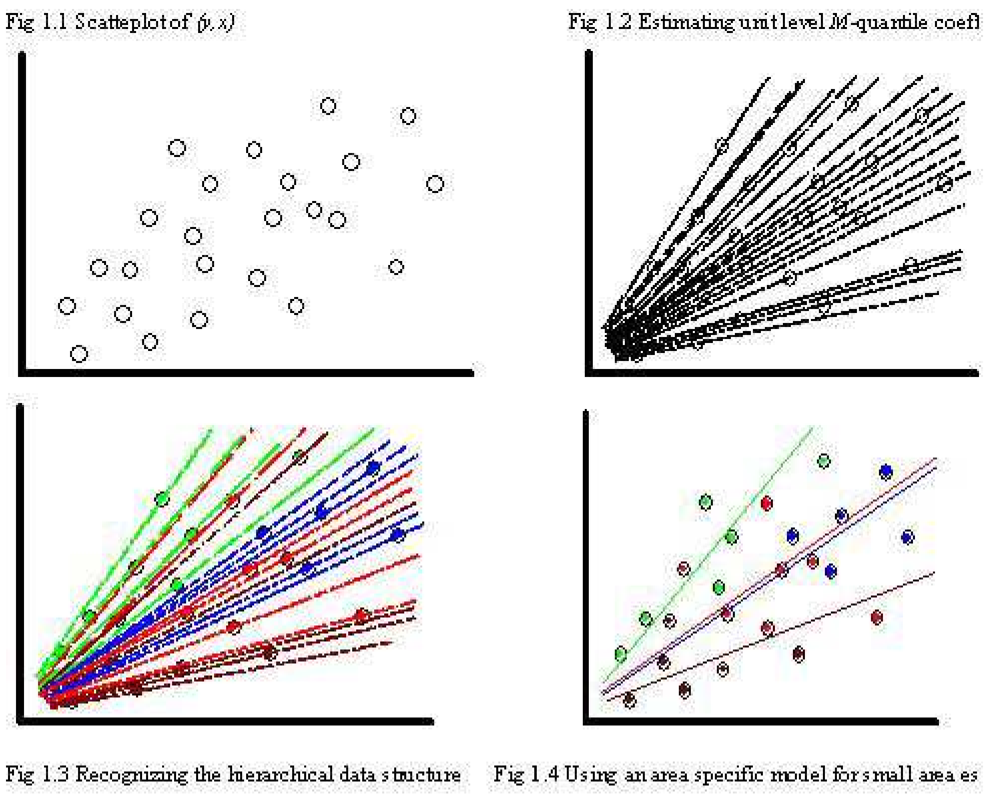
- accurately capture the relationship between *Y* and *X* in the data by exploiting this heterogeneity
- provide a decomposition of residual heterogeneity into between and within group heterogeneity, allowing an assessment of the impact of group structure on *Y* after allowing for *X*

An alternative approach to modelling the variability in this conditional distribution is via *M*-quantile regression, which does not depend on a hierarchical structure (Chambers and Tzavidis 2006). Let us assume that we have individual level sample data *(y,x)* on *Y* and *X* and an *M*-quantile model is used for modelling different quantiles *q* of the conditional distribution of Y given X, $m_q(x) = a(q) + \beta(q)X$. For fixed *x*, $m_q(x)$ is continuous in *q*, which implies that each sample value *(y,x)* will lie on one and only one regression *M*-quantile line. We refer to the *q*-value of this regression *M*-quantile as the *M*-quantile coefficient of the corresponding sample unit. By construction, the individual *M*-quantile coefficients *q* (which vary between 0 and 1) represent dimensionless measures of the residual heterogeneity in *Y* after heterogeneity in *X* has been conditioned away. We therefore have evidence of group related heterogeneity if the *q* coefficients tend to be more alike within groups than between groups.

In order to estimate the *M*-quantile coefficients we adopt the following approach. We first define a grid of *q*-values, e.g. *g* = (0.001, 0.002, 0.005…, 0.998, 0.999) that adequately "covers" the conditional distribution of *Y* and *X*. At the second step we fit a linear *M*-quantile regression model at each *q*-value in *g* and use linear interpolation to compute a unique *M*-quantile coefficient *q* for each individual *i* in the sample. We then compute a group level *M*-quantile coefficient by suitably averaging the *q* values of each sampled individual in that group. The final step involves fitting an *M*-quantile model at the group level *M*-quantile coefficient. This model is then use for estimating target parameters at the group (small area) level. The procedure is illustrated graphically in Fig.1.

During this project we also observed that M-quantile estimates of small area means were biased with the magnitude of the bias being related to the presence of outliers in the data. This led us to propose a bias adjustment to the M-quantile small area estimator of the mean that is based on representing this estimator as a functional of the small area distribution function (Tzavidis and Chambers 2005;2006). This approach was then generalized for estimating other quantiles of the distribution function in a small area.

Fig 1. *M*-quantile regression modelling of hierarchical data



Key Findings and Further Work

The project illustrated that an alternative approach to small area estimation is offered by *M*-quantile models not only for estimating small area averages but also for estimating other quantiles of the small area distribution function. *M*-quantile models open new areas for research in the use of mixed models. For example, the existence of hierarchical effects may be checked using diagnostic tests based on the *M*-quantile coefficients associated with this hierarchy. Alternatively, tests for the significance or otherwise of the coefficients of the *M*-quantile model may be useful in deciding on an appropriate formulation for the random part of a mixed effects model. This is an area of on-going research. An important problem in small area estimation is the impact of changing small area geographies on the estimates. It is difficult to resolve this using mixed effects models since the random effects in these models are geography-dependent, and therefore a change in the geography requires new random effects to be estimated. Such problems do not arise with *M*-quantile models since *M*-quantile coefficients are unit-specific and therefore do not change when a new geography is introduced. This allows the *M*-quantile coefficients for areas defined by the new geography to be obtained by simply reaggregating these unit-level M-quantile coefficients within these new areas.

Additional areas of current research involve

- Developing spatial *M*-quantile Models
- Estimating income distributions for UK Local Authority Districts
- Further evaluation of Mean Square Error estimators of the *M*-quantile small area estimates

Key Publications

Tzavidis, N. and Chambers, R. (2005): Bias Adjusted Small Area Estimation with M - quantile Models, Statistics in Transition, 7, 707 - 713

Chambers, R. and Tzavidis, N. (2006). M-quantile Models for Small Area Estimation, Biometrika, 93(2) 255-268

Working Papers

Tzavidis, N. and Chambers, R. (2006). Bias Adjusted Estimation for Small Areas with Outlying Values, Southampton Statistical Sciences Research Institute, 26pp. (S3RI Methodology Working Papers, M06/09) http://www.s3ri.soton.ac.uk/publications/methodology.php)

Further details are available from:
Ray Chambers
Email: ray@uow.edu.au
Phone Number: (02) 4221 3845