Research report
ESRC project R000238217
Applications and understandings of multilevel modelling in the social sciences

The aims of the project are:

1. An extension of software tools, to provide appropriate and informative ways for users to interact with their data in a flexible modelling environment.
2. Provision of workshops and training materials generally specially tailored for the needs of social scientists.
3. Involving users in extending the range of models that can be implemented. This is intended to provide models of sufficient sophistication to match the complexity of social science data.

Work proceeded smoothly on the development of methods and software tools. The new version of MLwiN (2.0 beta version, released Summer 2003) contains an enhanced interface, allowing the specification of simple, single level models with a smooth transition to multilevel models. It incorporates most of the project's methodological developments, most importantly a whole new set of models centred on MCMC (Bayesian) algorithms. These allow the user to fit a large range of multilevel models including: Measurement error models, large cross-classified and multiple membership models, missing data, complex covariance structures at level 1, factor analysis, and CAR spatial models. The project has focused on MCMC implementations for several reasons, including the ability to handle efficiently large cross-classified and multiple membership structures, and the provision of exact interval estimates.

Other major methodological developments are as follows. The application of a non-parametric (residuals) bootstrap procedure for multilevel models that allows some relaxation of model distributional assumptions; the use of weights at any level of a data hierarchy; efficient (simulated) maximum likelihood estimation for complex models; the use of model diagnostic procedures; the definition of a new measure of relative amounts of variation explained across the levels of a hierarchy; flexible models for event history data that allow a full multilevel specification of multiple state and competing risk models with general covariate structures.

Considerable attention has been paid to developing a formal diagrammatic representation of nesting and membership relations together with a simplified notation. With the increasing complexity of data structures being modelled this allows a relatively even the most complex structures to be described using a small number of basic building blocks. This can be done using simple classification diagrams and a special classification notation. The notation has been implemented in the software.

An interface has been built between *MLwiN* and WINBUGS, which is available in the new release version of the *MLwiN* software. A future area of research will be to explore re-engineering *MLwiN* using object-orientated, component based methodologies to increase *MLwiN*'s interoperability. That is, to make *MLwiN*'s estimation algorithms and user interfaces available to other software systems and vice-versa.

Training and dissemination has taken place using numerous 2-3 day workshops, seminars and conferences. The project has operated a monthly 'clinic' for researchers and provided email support. The multilevel modelling newsletter continues to be produced, as an electronic publication twice a year. It acts as a useful forum for

researchers to publish work in progress and to keep users in touch with developments. The Centre takes an active part in the multilevel email discussion group that it helped to found in the mid 1990s.

The project web site (http://multilevel.ioe.ac.uk) continues to be developed. It contains information about current activities, including coverage of all multilevel workshops and newsletters for downloading. It also contains current information about *MLwiN*. Building on the TRAMSS experience, the project has commissioned three researchers to produce on-line training materials in the areas of multilevel models for multiple category data in education, multilevel models for contextual and compositional effects in medical geography and public health. These will provide users with data and lead them through multilevel analyses using a freely available version of *MLwiN*. It is planned to develop this into a set of materials that covers all the major social science areas of application. A series of reviews of major multilevel software packages has been undertaken and is published on the web site.

An important part of the project's work has been collaboration with other users and researchers. The project has been able to collaborate with all those mentioned in the grant application. In addition several individuals have contributed to developments through taking part in seminars, joint data analyses or the regular monthly meetings organised by the Centre for Multilevel Modelling for associated 'fellows'. These include Alice McLeod, Alastair Leyland, Anthony Fielding, Danny Pfeffermann, David Spiegelhalter, Dougal Hutchison, Edmond Ng, Gad Nathan (long term visitor), Gilbert Mackenzie, Ian Schagen, Kelvyn Jones, James Carpenter, John Nelder, Michael Healy, Nicky Best, Nigel Rice, Paul Bassett, Sylvia Richardson, Toby Lewis, Vanessa Simonite, and Youngjo Lee.

In addition, the project has continued to develop institutional collaborations, particularly with government departments such as ONS and DfES in the UK and the ministry of Education in France. International collaborators include researchers drawn from the fields of genetics, epidemiology, demography, HIV/AIDS research, education and political science.

Many of the collaborations have resulted in publications in discipline areas including the following: a collaboration with Michel Poulain resulting in the use of multiple membership models to fit extensive Belgian longitudinal household data; a collaboration with Tom O'Connor and Jenny Jenkins on the application of multilevel models to cross-sectional and longitudinal family data; a collaboration with Richard Pettifor applying multilevel models to evolutionary ecology; a collaboration with Anthony Heath on an analysis of longitudinal election study data; a collaboration with Risto Lehtonen developing multilevel models for finite population inference using an extension of generalized regression estimation.