

Award no. R000222732

Award Holders: Prof. H. Goldstein, J. Rasbash

Title: Developing graphical and inferential tools for social data analysis

Summary of Research Results

Final report; 1000 word summary

In recent years, due partly to the increasing role of the computer, methodology and software have become available which allow researchers to study statistical models with increasingly complex data structures. The ESRC ALCD programme has helped with the development of both MCMC and likelihood based methods for fitting multilevel models to these complex data structures, in part through their support of the computer software packages BUGS and *MLwiN*.

Both of these software packages have increased the group of researchers that can fit complex statistical models from a small group of statistical methodologists to the social and medical science research community in general. This increase in practitioners of multilevel analyses has thrown up models of increasingly higher levels of complexity, and has led to new insights in a number of areas such as education, geography, demography and political science. The increase in interest is leading to models of greater complexity and different statistical estimation methods to fit these new models.

The first version of the software package *MLwiN* released to the user community in February 1998 included both maximum likelihood and Bayesian MCMC estimation methods for fitting multilevel models (Browne 1998), and also included another simulation-based technique, bootstrapping (Efron and Tibshirani 1993) to fit multilevel models in *MLwiN*. It was the first package to incorporate all these features and raised the issue of how users could make sensible choices about which to use.

This research project had 3 main aims. The first was to provide guidance to users on the use of these techniques using the graphical user interface in *MLwiN* together with any necessary enhancements. The project has carried out simulation studies to compare the various estimation methods available for fitting multilevel models, concentrating on comparing the likelihood-based methods and the various Bayesian MCMC methods for a few of the most common multilevel models. Following the simulation studies and discussion with the user community through both workshops and meetings modifications and enhancements have been made to many of the features of the graphical interface, and these are being incorporated into the second major release of *MLwiN* in late 1999.

The second aim was to provide a body of exemplar applications to guide users. The manual that accompanied the first release of *MLwiN* contained only two short chapters on the use of the simulation-based methods. During the course of this project an updated version of the user manual has been written to complement the second major release of the software. This new manual both introduces the new features in the software and expands on the existing material on the simulation methods. The new manual has three sections, with the third section solely concerned with the MCMC and bootstrap methods. This section contains an introductory chapter on simulation-based methods to motivate new users in their use. There are then 3 chapters on the MCMC methods which include more of the theory surrounding the method as well as greater explanation of the various MCMC methods and options in the software package. The bootstrapping chapter has also been expanded to include a new non-parametric bootstrapping method based on sampling from modified residuals.

The final aim was to disseminate knowledge about these techniques through workshops, conferences and via the internet. During the course of the project there have been 6 *MLwiN* workshops which have incorporated the simulation based methods, including 1 workshop based

solely on the MCMC options in the package. This provided an opportunity to show new users that there are alternative methods, both classical and Bayesian, for fitting multilevel models and to explain to them a little of the theory and motivation behind these methods. These workshops were not only a good means of dissemination for the project but also were good opportunities for getting opinions from the user community.

During the course of the year Dr Browne and Professor David Draper, who acted as a consultant to the project, gave talks at the 2nd International Amsterdam conference on Multilevel Analysis based on two forthcoming joint papers that were well received. Talks were also given to both the health economics group at the University of York and the MRC social and public health sciences unit at the University of Glasgow on MCMC estimation methods for fitting multilevel models. Before the project was started there was an existing web site for the *MLwiN* package (<http://www.ioe.ac.uk/MLwiN>) which contained details of the software and was used to inform users of developments as well as allowing them to download upgrades of the software. The project expanded the site with 'bug' reporting pages which report known problems in the package and in which version they are fixed. During the pre-Beta testing period of the second release of *MLwiN* (version 1.1) the project has created a web site to enable the researchers who are testing the software to download the latest versions of the software, manual chapters and help system. This web site also includes a discussion list to allow testers to post problems and for us to announce new improvements. It is anticipated that this web site will be a template for a full Beta release web site in October/November 1999.