**Generalised Structure Preserving (GSPREE) Models in Small Area Estimation Application in estimation of ethnic group size for local authorities**

Philip Clarke; Alison Whitworth; Office for National Statistics

The Office for National Statistics is looking at using more administrative and survey data to produce typical census outputs, in an Administrative Data Census. However, survey data often have very small or null sample sizes within areas, and administrative data may not cover the entire population. The census will continue to be a detailed source of information for some time after it is carried out, but for intercensal estimates it would also become outdated. This leads to the use of data sources in combination to produce reliable estimates for small areas and sub-groups of the population for which there are small or no samples. Small Area Estimation (SAE) methods provide a framework for combining data sources in Official Statistics.

Many cases of SAE involve continuous variables such as income or simple binary categories such as poverty. For these multilevel regression models using correlating administrative and census variables are in common use. However, where the interest is in multiple categories and where particular correlating variables are harder to justify, then these methods are less flexible.

Generalised Structure Preserving Estimation (GSPREE) can combine low precision cross classified small area by variable-of-interest data (e.g. from sample survey) with higher precision data of the same cross classification (e.g. from administrative data) and with aggregate margin data in a model to produce good estimates for the categorical variable-of-interest. Such examples are population size by ethnic group or by household characteristics. The methods are flexible in that additional data sources can be introduced as available and existing sources excluded if they are no longer relevant.

GSPREE has been used to produce population estimates for ethnic groups by local authorities in England (Zhang and Chambers, 2004; Luna-Hernandez et al., 2015). The performance of the GSPREE estimator has now been assessed in a validation scenario where the true population distribution is known (i.e. March 2011 Census). A GSPREE model considered the contribution of different auxiliary information (2001 Census and 2011 English School Census) for various age groups, alongside survey data. Uncertainty in the GSPREE estimates was estimated using a bootstrap.

Key Words: Small Area Estimation, GSPREE, admin data, combining data, ethnicity

Contact Details for Philip Clarke: Contact Details for Alison Whitworth:

Telephone: 02075928647 Telephone: 01329444653

Email: [philip.clarke@ons.gsi.gov.uk](mailto:philip.clarke@ons.gsi.gov.uk) Email: [alison.whitworth@ons.gsi.gov.uk](mailto:alison.whitworth@ons.gsi.gov.uk)