

PAPER TITLE: SPATIALMSM: The Australian spatial microsimulation model

Robert Tanton
National Centre for Social and Economic Modelling
University of Canberra
Australian Capital Territory 2601
Australia
Robert.tanton@natsem.canberra.edu.au

ABSTRACT

In recent years, the National Centre for Social and Economic Modelling at the University of Canberra has been developing a method of calculating estimates for small areas using survey data. This methodology has now been linked to our microsimulation model of the Tax/Transfer system STINMOD, allowing the regional impact of policy changes to be estimated.

The SPATIALMSM model reweights the survey data to small area known reliable totals (benchmarks) from a Census. There are a number of benchmarks used, and these depend on the survey variable being estimated. For estimating poverty and housing stress, we use labour force status, age, sex, income, rent paid, mortgage paid, household type, Household size (number of people per household), dwelling tenure, dwelling structure, Non Private Dwelling, and total number of households. All these variables are correlated with poverty and housing stress, so help to provide the best estimates of these two variables.

The weights for each small area are then applied to variables from the survey (equivalised disposable income for both poverty and housing stress, and housing costs for housing stress), and small area estimates for these variables can be calculated. The accuracy of these small area estimates depends on the correlation between the variable being estimated and the benchmarks used, so the benchmarks must be chosen with the final variable to be estimated in mind.

This method is now being used in two grants to calculate small area estimates of needs based planning indicators (for instance, number of people aged 65 and over; number of people receiving rent assistance), and for estimating variables affecting children (for instance, children in households in poverty; children in overcrowded households). In future, we expect to derive small area estimates of variables affecting older people.

OBJECTIVE OF THE MODEL

The main objective of the model is to allow the calculation of small area estimates of survey data. The main problem with survey data is that while it provides a huge number of variables, by necessity, the sample sizes are designed to allow reliable estimates for large areas, but not for smaller areas. For smaller areas, sample sizes mean that estimates are unreliable; and in out of sample areas, non-existent.

A second objective of the SpatialMSM model is to allow the effect of policy simulations at a small area level to be examined. This is done by linking NATSEM's STINMOD microsimulation model to the SpatialMSM regional microsimulation model. Both models use the same base data (the 2002-03 Survey of Income and Housing Costs and the 2003-04 Survey of Income and Housing), allowing the linking of unit records from STINMOD to unit records from the SpatialMSM model.

Because the model is linked to our STINMOD microsimulation model, this also means projections can be derived, using STINMOD's outyears capability. To provide an outyears capability for SpatialMSM, the weights calculated by the SpatialMSM model for each small area are inflated by the population projections for each projection year. The weights for each projection year are then linked to the STINMOD outyears file for that year to derive small area projections.

TARGET AUDIENCE

There are a number of target audiences for the SpatialMSM model. The main target audience for the model is service delivery agencies. These are agencies in each State who are delivering services to small areas. In Australia, they include local government (mainly interested in land use and potential demand for housing), State government (providing education, health and aged services) and the Commonwealth government (providing funding for the State government expenditure on many services). All these levels of government are interested in information on their clients in small areas where they provide services.

Another target audience are researchers, who want to analyse relationships at the smallest area possible to minimise any averaging that may be in an area. There is also a lot of interest in Australia in defining 'neighbourhoods', and to identify a neighbourhood, the smallest level of geography is required so aggregations existing in the community can be reflected in the data.

The third target audience is policy makers in Australian government. Policy makers are interested in estimating the effect of a policy change for small areas, so they can see the spatial effect of any policy change. Linking the SpatialMSM model with NATSEM's model of the Tax/Transfer system (STINMOD) has meant we can estimate policy changes at a spatial level.

Sources of small area data to assist these target audiences are available from the ABS Census or administrative data. Both sources provide a very limited set of information – so, for instance, administrative Taxation data will give information on individual taxable income in an area; but it will not give equivalised disposable household income. The ABS Census provides demographic information for small areas; but it is only conducted once every five years, so it becomes out of date.

The other option available to these audiences for information is survey data. However, in any survey, there is a trade off between sample size and cost. Increasing the sample size means more reliable data for smaller areas; but also increases the cost. The sample size for most Australian surveys conducted by the Australian Bureau of Statistics (ABS) are designed to provide reliable estimates for Capital City/Rest of State in each Australian State; but no smaller than this. The ABS doesn't provide geographic areas below Major Urban/Other Urban/Other (called Section of State) on the Confidentialised Unit Record Files (CURFs – these are record unit data from the ABS surveys, confidentialised to eliminate the possibility of identifying anyone in the survey).

Because of the lack of detailed information for small geographic areas in Australia, there has been a lot of interest in creating synthetic estimates for these small areas. The ABS recently published a paper reviewing the methods for small area estimation, and providing guidance on diagnostics for small area estimates (ABS, 2006). A number of Commonwealth Government agencies also use small area methods to calculate estimates for small areas. Estimates available in Australia include estimates of disability (Elazar and Conn, 2005), employment (Commonwealth Department of Employment and Workplace Relations, 2007) and crime (Tanton, *et al.*, 2001).

METHOD AND PLATFORM

The SpatialMSM model uses the reweighting SAS macro called GREGWT. This was written by the ABS to enable reweighting their survey data to reliable benchmarks available from other sources (the Estimated Resident Population, ERP). A description of the macro can be found in Bell (2000).

The GREGWT algorithm is essentially a constrained distance minimisation function. The method uses regression to get an initial weight; and then iterates the regression until convergence is reached (so the difference between the estimated benchmark and the actual benchmark for the area from the Census data is within a set limit), or a set number of iterations is made, by which time the iteration stops. The process needs a start weight, and this is set to the original ABS survey weight for the survey record divided by the population of the area being estimated. In many cases, there is no iteration as the initial regression estimate provides weights that are within the tolerance set.

Because the macro is programmed in SAS, the SpatialMSM model is also programmed in SAS. This also means it is fairly easy to link into STINMOD, as STINMOD is programmed in SAS.

We have also recently developed an Excel front end, that allows any of the values for each benchmark to be changed, and the reweighting conducted again to derive new estimates using the revised values.

POLICY ENVIRONMENT WHERE IT HAS BEEN USED

So far, the SpatialMSM model has been used to provide estimates of before and after housing poverty (to be published by the Australian Housing and Urban Research Institute), and regional estimates of poverty (Chin 2006) and housing stress (Kelly, *et al.*, 2005, Phillips, *et al.*, 2006). The paper by Kelly, *et al.* (2005) also derived

projections of housing stress. A paper on poverty and inequality was published by Harding *et al* (Harding, *et al.*, 2006)

The model has also been used for policy simulations, to simulate regionally the effect of the tax cuts made by the Australian Government in 2005. This work was published in Chin, *et al.* (2005).

We are currently working on two grants using this methodology, one to estimate service delivery populations with a number of Australian State partners (ARC LP0775396); and one to estimate data for young people and old people (ARC DP0664429).

REFERENCES TO TECHNICAL PAPERS

There have been a number of technical papers on the SpatialMSM model. The first technical paper from NATSEM outlines the method used for manipulating the survey data that links to the Census data (Chin, *et al.*, 2006). The second technical paper outlines the reweighting methodology (Chin and Harding, 2006).

A brief summary of how the model works can be found in Chin and Harding (2007).

BIBLIOGRAPHY

- ABS (2006) *A guide to small area estimation - Version 1.0*,
[http://www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/3a60738d0abdf98cca2571ab00242664/\\$FILE/May%2006.pdf](http://www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/3a60738d0abdf98cca2571ab00242664/$FILE/May%2006.pdf), Last accessed 30 July 2007
- Bell, P (2000) *GREGWT and TABLE macros - Users guide*, Canberra: ABS, Pub. # Unpublished
- Chin, S-F and Harding, A (2006) *Regional Dimensions: Creating synthetic small-area microdata and spatial microsimulation models*, Technical Paper 33, NATSEM
- Chin, S-F, Harding, A and Bill, A (2006) *Regional dimensions: Preparation of 1998-99 HES for reweighting to small area benchmarks*, Technical Paper 34, NATSEM
- Chin, S-F, Harding, A and Tanton, R (2006) 'A spatial portrait of disadvantage: Income poverty by Statistical Local Area in 2001', Paper given at 2006 ANZRSI Conference Heritage and regional development, Beechworth, Victoria, 26 - 29 September 2006
- Chin, SF and Harding, A 2007, 'SpatialMSM - NATSEM's small area household model for Australia' in A. Gupta and A. Harding, (eds), *Modelling our future: Population ageing health and aged care*.Oxford, Elsevier, vol. pp. 563 - 566.
- Chin, SF, Harding, A, Lloyd, R, McNamara, J, Phillips, B and Quoc, NV (2005) 'Spatial Microsimulation Using Synthetic Small-area estimates of Income, Tax and Social Security Benefits', *Australasian Journal of Regional Studies*, Vol. 11(No. 3), pp. 303 - 335
- Commonwealth Department of Employment and Workplace Relations (2007) *Small Area Labour Markets*, Commonwealth Government
- Elazar, D and Conn, L (2005) *Small area estimates of Disability in Australia*, Canberra: ABS, Pub. # 1351.0.55.006
- Harding, A, Lloyd, R, Bill, A and King, A 2006, 'Assessing poverty and inequality at a detailed regional level: New advances in spatial microsimulation' in M. McGillivray and M. Clarke, (eds), *Understanding Human Well-being*.Helsinki, United Nations University Press, vol. pp. 239 - 261.
- Kelly, S, Phillips, B and Taylor, E (2005) *Baseline small area projections of the demand for housing assistance*, Final Report, AHURI
- Phillips, B, Chin, SF and Harding, A (2006) 'Housing Stress Today: Estimates for Statistical Local Areas in 2005', Paper given at Australian Consortium for Social and Political Research Incorporated Conference, Sydney, 10-13 December 2006
- Tanton, R, Jones, R and Lubulwa, G (2001) 'Analyses of the 1998 Australian National Crime and Safety Survey', Paper given at The Character, Impact and Prevention of Crime in Regional Australia, Townsville, 2 - 3 August 2001