



Pushing them to the edge

An assessment of spatial microsimulation

Robert Tanton
Yogi Vidyattama

IMA Conference, Ottawa, June 2009

Acknowledgements

- Funding for this project was provided by an Australian Research Council Linkage Grant (LP0775396). The partners on this grant included the NSW Department of Community Services; the Australian Bureau of Statistics; the ACT Chief Minister's Department; the Queensland Department of Premier and Cabinet; Queensland Treasury; and the Victorian Departments of Education and Early Childhood and Planning and Community Development.

Structure

- Data
- Methods
- Results
- Conclusions

Data

- 2002/03 and 2003/04 ABS Survey of Income and Housing
- 2006 Census for small area benchmarks
- Statistical Local Areas used
 - No problems with confidentiality
 - Covers whole of Australia

Method

- Benchmarks from Census

Number	Benchmark
1	Age by sex by labour force status
2	Total number of households by dwelling type (Occupied private dwelling/Non private dwelling)
3	Tenure by weekly household rent
4	Tenure by household type
5	Dwelling structure by household family composition
6	Number of adults usually resident in household
7	Number of children usually resident in household
8	Monthly household mortgage by weekly household income
9	Persons in non-private dwellings
10	Tenure type by weekly household income
11	Weekly household rent by weekly household income

Method

- Reweight survey data to small area Census benchmarks
- Constrained optimization technique
 - Uses Chi Squared distance function constrained to benchmarks
 - deterministic
 - Additional constraint added by us because we don't want weights less than 0
 - Iterative – maximum iterations set to 30 – if can't find a solution by then, won't converge.

Covergence

- Find sometimes when won't converge, estimates can still be used
- Added Total Absolute Error criteria
 - If sum of absolute value for all errors less than total population in small area, then can use.
- Only 0.7 per cent of total population in areas with failed TAE
- 1284 SLAs out of 1422 converged

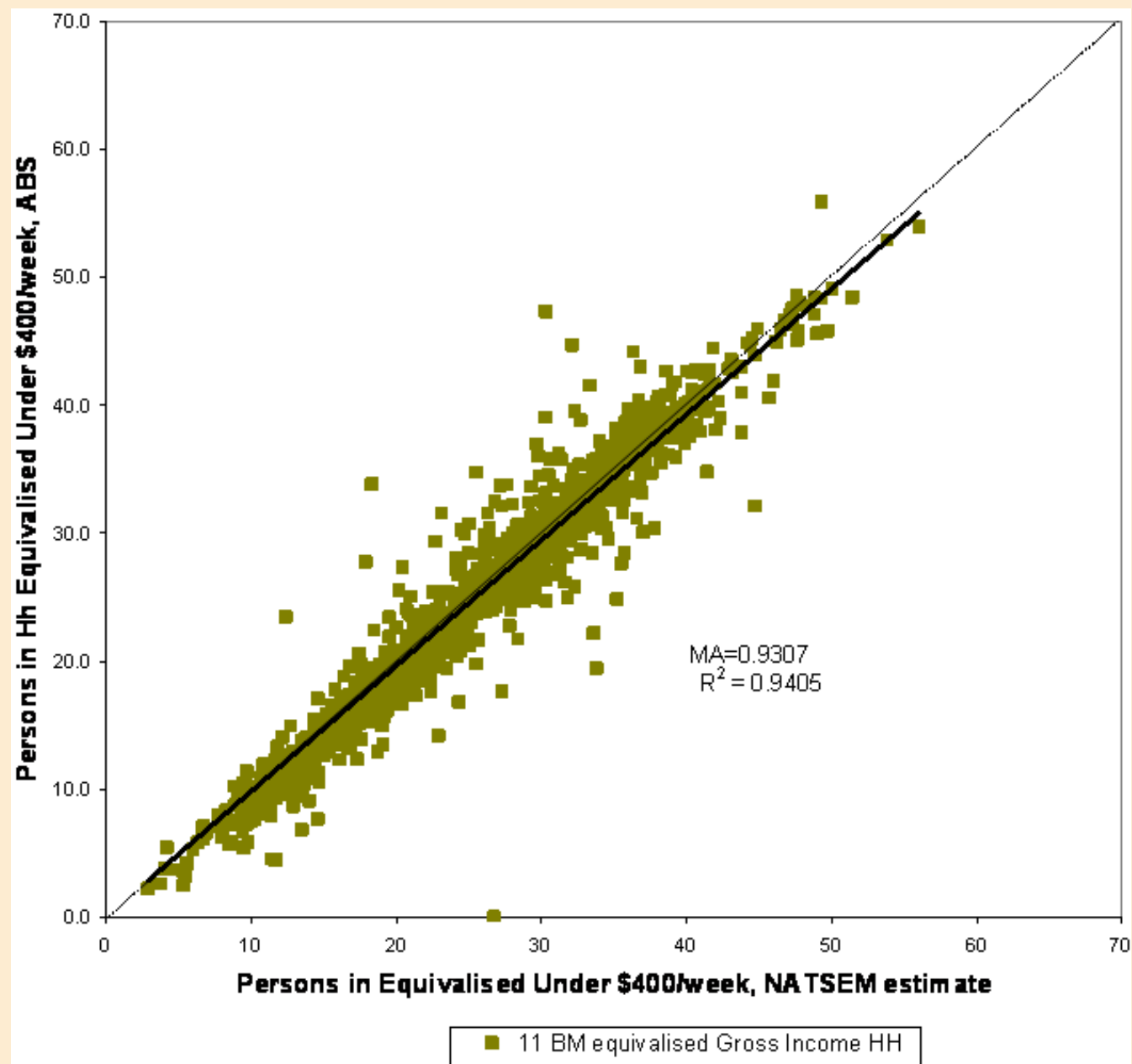
Measure of accuracy

- Once passed TAE test, how good are our estimates?
- Compare to *unconstrained* estimates from Census for small areas
 - Unconstrained means not in our constraint tables – should get very good estimates for constrained variables.
 - Need to simplify estimate because Census has very limited information
- Proportion of people in poverty with poverty line set at \$400 gross income
 - Match using SpatialMSM

Measure of Accuracy

- Plot of small area Census data against modelled data
- Expect 45 degree line – error is how far we are off this line, measured using Measure of Accuracy

$$MA = 1 - \frac{\sum (y_{est} - y_{ABS})^2}{\sum (y_{ABS} - \bar{y}_{ABS})^2}$$



Changes made – Add benchmarks

- Add benchmarks
 - Expect greater non-convergence because model is more complicated
 - Expect greater measure of accuracy for some outputs
- Non-school qualifications
- Occupation

Changes made – Univariate benchmarks

- Derive greater convergence using simpler benchmarks
- Use univariate rather than multivariate benchmarks
- Expect lower accuracy
- End up with 14 benchmarks rather than 11

Changes made – Limiting the source of households for the estimation

- How well do survey observations from outside Sydney represent people in Sydney? And vice versa
- Run model for Sydney with Australian sample; and Sydney only sample
- Repeat for other capital cities in Australia

Results – Adding additional benchmarks

- Additional 2 benchmarks reduced number of SLAs from 1284 to 1257
- Measure of accuracy increased slightly from 0.9307 to 0.9388

Results – Univariate benchmarks

- As expected, number of converging SLAs increased from 1284 to 1329
- However measure of accuracy decreased from 0.9307 to 0.8781
- Partly to do with including SLAs which even though TAE showed reasonable estimates against constrained variable, did not give good estimates of unconstrained variable
- If use SLAs passed TAE test from base model, get MA = 0.91 – so still below base model MA of 0.9307

Limiting the source of households from the sample

- Very little effect on the number of SLAs passing TAE test
- Very little effect on Measure of Accuracy
 - Greatest effect for Perth – better estimates when using Perth sample rather than Australian sample, also same for Adelaide but not as great
 - Melbourne gave better estimate using Australian sample - only city that did
 - Sydney and Brisbane very little difference

Conclusions

- SpatialMSM model very robust to changes – very stable
- Can add benchmarks with no great decrease in measure of accuracy and slight increase in number of SLAs failing TAE criteria
- Best results using multivariate benchmarks – univariate benchmarks give more SLAs passing TAE criteria but at the expense of accuracy
- Using all households in model has small effect on smaller capital cities in Australia, but no significant effect on larger capital cities and no effect on number of SLAs passing TAE criteria



3283 2523 7564 4857
7564 2734 3165 2791
0288 3534 8110 3372

www.natsem.canberra.edu.au