Small Area Coupling of Synthetic Census and Spatial Micro-simulation Applied to Mortgage Taxation in Australia

R. Tanton [a], **P. Perez** [b], **C. Pettit** [c], **Y. Vidyattama** [a], **J. Li** [a], **S. Zarpelon Leao** [c]

[a] *National Centre for Social and Economic Modelling, IGPA, University of Canberra*
[b] *Smart Infrastructure Centre, University of Wollongong*
[c] *City Futures Research Centre, University of New South Wales*

Email: robert.tanton@canberra.edu.au

Abstract: Socially disaggregated and spatially explicit models, like spatial micro-simulation or agent-based modelling, allow for flexible testing of policy interventions at household level and analyzing of their consequences at relevant levels of socio-economic or spatial aggregation. However, the approach requires demographic data both at household and small area levels, information that is not publicly available in Australia. The authors propose to use a synthetic census to generate this information and couple it with a spatial micro-simulator. This paper describes the approach and provides an example of application to taxation in Australia. The application examines the impact of making mortgage interest payments tax deductible, similar to the way that investment property mortgage payments are tax deductible.

Results show that, overall, the proposed change in taxation regime was nearly cost neutral by increasing tax rates for top income earners. The impact of this change is looked at by family type, household income decile and the area that the person is living in. the model shows that this housing policy benefits low income households most, and the impact is relatively similar across all Australian States and Territories. Households with a household head aged 16 – 44 benefit more than those with an older household head, although households who lose most have a household head aged 35 – 44. Single parents with dependent children win most, and couples with dependent children lose most.

These results also show that households in regional and remote areas were mainly in the middle of the distribution in terms of losses, but experienced significant gains due to the rental policy change. Most households in richer areas lost, while households in poorer areas won, due to a mix of lower incomes and greater benefits from being able to claim mortgage payments as a deduction against their income.

From a modelling perspective, the study shows that bringing together a synthetic census and a spatial micro-simulation model is possible using direct linking of the household ID, as long as the two models use the same base dataset. Where this is not the case, then other methods can be used. However, these methods will not provide the same amount of variability as a direct linking.

Keywords: *Synthetic population, micro-simulation, mortgage, small area estimation*

1. INTRODUCTION

Housing is an important part of the Australian dream which is becoming harder to achieve for many young Australians. The latest 2016 Australian Census showed that a much lower proportion of people own their house in 2016, and a higher proportion are renting or purchasing. Analysis of the Census data by age group shows that it is mainly younger and middle age families (age 25 – 44) who are affected by this decline in home ownership, with older families (age 55 and above) experiencing little or no decline in home ownership rates from 1981 to 2016. The reason for this is mainly due to house prices increasing in Australian capital cities much faster than incomes. In 1981, the median mortgage for 25-34-year-olds was 17 per cent of household income, but by 2011 it was 25 per cent. Over this period, the income to house price ratio (how many years of median income is required to purchase a median priced house) has increased from 3.3 years in 1981 to over 7 years (Department of Parliamentary Library 2016). Recent research looking at generational change in home ownership by members of AHURI shows that from 1981 to 2011 there was a decline in home ownership from 61.4% to 48.4% for 25–34 year-olds and from 74.3% to 65.3% for 35–44 year-olds. Much of this decline was from 1981–91, and the change since then has been lower. The authors attribute some of this decline to the changing nature of mortgages, where a mortgage can now be used as a cheap line of credit for renovations, to buy a car, or for an overseas trip, so there is little incentive for fully paying off a mortgage (Burke et al. 2014).

The tax and transfer system is one policy lever controlled by the Commonwealth government that might be able to assist younger families with lower incomes get into the housing market. This could be through financial encouragement to purchase a principal residence, either through a direct grant (like the current State Government-funded First Home Owners Grant) or through financial incentives paid through the tax and transfer system. One option under the tax/transfer system is to make interest payments for the principal residence tax free, as they are for investment properties. In this modelling, we have allowed a tax deduction on interest payments for the principal place of residence of up to $200 per fortnight. In any change to the tax/transfer system, when there are increased benefits for a group, there are increased losses to another group to offset these gains, so that the overall impact is budget neutral. For this analysis, we have funded the change by increasing taxes for high income earners (those earning more than 180,000) from 37c in the dollar to 49c; and the second highest tax rate (for those earning between $87,001 - $180,000) from 32.5c in the dollar to 38.5c. These increases balanced out the additional expenditure of the mortgage payment tax deduction.

Modelling these changes to the tax/transfer system requires a socially disaggregated model that operates at household level (Li et al. 2014). Furthermore, as impacts of housing policy are geographically localized, the approach requires a spatially explicit model (Robert Tanton and Clarke 2014). These two modelling paradigms have been brought together previously for studying the spatial impact of a policy change (Harding et al. 2009), and this paper extends this work by joining a tax/transfer microsimulation model with a synthetic census for Australia. The advantage of this approach is that other household based models can be linked in the future, as suggested in recent papers by the authors (Robert Tanton, Perez, et al. 2017).

Spatial microsimulation models can use reweighting methods or a synthetic reconstruction approach. Previous work by the some of the authors of this paper have used a reweighting approach (Robert Tanton et al. 2011), but more recently to allow linking of agent based models, a synthetic reconstruction approach has been used (Harland 2013; Huynh et al., 2016). While there are limitations to this approach, including the fact that the current implementation cannot include both household and person level benchmarks (Robert Tanton, Huynh, et al. 2017), the synthetic population created can be used as a proof of concept on how to link a tax/transfer microsimulation model and a synthetic population.

While a tax/transfer model has already been linked to a reweighted small area population model (Harding et al. 2009), it has not been done for a full synthetic population. This paper demonstrates that the approach suggested in Tanton, Perez, et al. (2017) is feasible, and that a synthetic population can be used as the core of a spatial decision support system using a number of integrated, already developed and tested, models. This paper also looks at how this could be implemented to improve decision making in cities by incorporating the modelling into the University of NSW IT systems being developed.

Although this paper focuses on the coupling of a synthetic census with a specific spatial microsimulation model, the approach can be generalized to other disaggregated modelling paradigms like agent-based modelling. Huynh and colleagues (2015) have coupled a synthetic reconstruction approach and an agent-based model of transport and residential mobility in South East Sydney. The model not only allows to potentially allocating households and individuals at street block level but also to evolve the synthetic population with a reasonable accuracy (Leao et al., 2017).

2. DATA

The microsimulation models required for this project need two sets of data. The first set is a unit record file for a national survey, which is used by both the tax/transfer microsimulation model and the spatial microsimulation model. The spatial microsimulation model derives a synthetic census of households across the whole of

Australia at the SA2 level, so also needs small area benchmarks at the SA2 level from the latest Australian Census (currently 2011). The SA2 geography was used because it is detailed enough in capital cities of Australia, while also being large enough in rural and remote areas. In capital cities, SA2's generally match to suburbs, so they are an area recognised by the local community. There are a total of 2,110 SA2's across Australia, but some of these are termed "Offshore and Migratory" (one for each State and Territory), and results are not derived for these areas, nor for industrial areas or non-residential areas. The population of SA2's ranges from 0 to 14,260 families with an average of 3,973 families per SA2.

3.   METHOD

3.1.   Tax/Transfer Microsimulation

Microsimulation models were first described by Guy Orcutt (Orcutt 1957), and have been developed extensively since then. The tax/transfer system is an ideal system for a microsimulation model as it is rules based, so the microsimulation model applies the rules from the tax/transfer system as they are specified by the Government (Li et al. 2014). In Australia, the Federal government derives the rules based on annual budgets.

The tax/transfer microsimulation model used for this paper was STINMOD+, a model of the Australian tax/transfer system design by the National Centre for Social and Economic Modelling (NATSEM) at the University of Canberra. STINMOD+ is a tax microsimulation model that can be used to evaluate the distributional and fiscal impact of tax and transfer policies. The model simulates the likely benefit amount and the tax liability of a household conditional on income and other socio-economic characteristics. The simulation covers all major taxation and welfare schemes (including subsidies and rebates) in Australia, and incorporates comprehensive income and asset testing (Li and La, 2017). The current version of STINMOD+ uses the 2011/12 and 2013/14 ABS Survey of Income and Housing as the base dataset and applies the rules for the Australian tax/transfer system as of July 2017.

The STINMOD+ model estimates the impact of the tax and transfer system at an income unit level. For this analysis, we have aggregated this to household level. Because the model is based on surveys, there are also weights which reweight the survey to the national total. In STINMOD+, these weights are at the income unit level, rather than the household level. For this analysis, we have used the original household level weights on the survey, and divided these by 2 as there are 2 surveys. These household weights will be similar to household weights derived from STINMOD+ if STINMOD+ had been calculated at a household level, rather than income unit level.

3.2.   Spatial Microsimulation

The spatial microsimulation method used is a combinatorial optimisation method with simulated annealing, as described in Harland (2013). The method is described in full in Williamson (2007). The method randomly selects the number of records required for an area (based on the small area Census data) from the base survey, and then randomly selects a household to remove and replace with another household from the survey. If the fit for the area improves, then the household is left in; if the fit gets worse, then the old household is put back in, and another household to replace is selected. This process continues until a set accuracy criteria is met, or a maximum number of iterations is reached. A simulated annealing process is also used so that the procedure can move back from a local maximum to derive an overall better result. This process is fully described in Williamson (2007) and Tanton (2014).

3.3.   Bringing the two models together

Previous work has shown how tax/transfer and a reweighted spatial microsimulation model can be linked (R Tanton et al. 2009). This method relies on the fact that the two models (the tax/transfer model and the spatial microsimulation model) had the same survey as the underlying base data. A reweighting spatial microsimulation model provides a reweighted unit record file for every area, so this can be merged with the unit record file from the tax/transfer microsimulation model, and the changes calculated for each small area. This is a one to one merge – there will be one household in the tax/transfer file to match with one household in each small area.

A synthetic population is different in that there are 8.8 million records for households across Australia, and by necessity, a survey of 14,569 households will need each household to be repeated many times. There are two options in terms of applying the identified change in the tax/transfer model to the synthetic population.

One method has been used by some of the authors of this paper when linking a CGE model to a tax/transfer model (Vidyattama et al. 2014). This method takes the change in income by industry (and occupation) modelled by the CGE model, as well as changes in the number of employed persons in an industry (and occupation) modelled by the CGE model, and applies these changes to the tax/transfer model base file. If the CGE model

calculates that a 2 per cent of people in retail industry of a State become unemployed, then 2 per cent of people in retail industry of that State in the tax/transfer model will be moved from employment to unemployment, and will be given the unemployment benefit calculated for their family type, number of dependents, and other criteria. The implementation of affected employment is much more difficult as this may involve moving a person from one sector of industry to another, which may involve the use of the person current salary, education and other individual characteristic (Vidyattama et al.,2014).

The other method is similar to the method used in Tanton et al. (2009), since the synthetic population uses the same base survey data as the STINMOD+ Tax/Transfer microsimulation model. This method merges the two models by the household ID, as both models are based on the unit records. One complication with this for our work is that the STINMOD+ model used two surveys (the 2011-12 and 2013-14 surveys), whereas the synthetic population only used one (the 2011-12 survey). This means we don't have all the families from the STINMOD+ model, but have enough to demonstrate the reliability of the method.

The tax/transfer microsimulation model STINMOD+ allows distributional results to be calculated, so results (winners and losers from the policy change) can be analysed by State, by family type, by age of oldest person in the household and by household income decile. These results do not need to be taken from the synthetic population, however results derived from the synthetic population should be the same as results derived from STINMOD+ as STINMOD+ is benchmarked to a similar set of variables at the national level as the synthetic census, so the aggregate State and Territory results should be similar the two models (but not exactly the same because STINMOD+ uses 2017 benchmarks while the synthetic population uses 2011 Census small area benchmarks). This paper tests whether this is the case. Using the synthetic population also allows winners and losers to be identified for each SA2 in Australia, so spatial results from the policy change can be identified.

The income deciles we have used for this analysis are equivalised household income deciles. Equivalised household income recognises that incomes are shared across all members of a household, but the economies of scale means that each additional household member does not add one unit of expenditure. The equivalising factor used is the OECD factor, which uses 1 point for the first adult; then 0.5 points for each additional person in the household aged 15 and over; and 0.3 for each additional person aged under 15.

4. RESULTS

4.1. Results by Age

The results by the age group of the oldest person in the household are shown in Figure 1. It can be seen that STINMOD+ and the synthetic population showed similar changes, with some differences for winners aged 25-34; 55-64; and 65+. There was not much difference in the results for losers. Generally, the synthetic population was below STINMOD+, although for aged 25-34 it was above. Generally the winners were the same across the different population groups, with about 2.3% of income, although for the oldest age group, the winners gained 2.8%. In terms of losers, the age groups 35-44 and 55-64 lost the most.
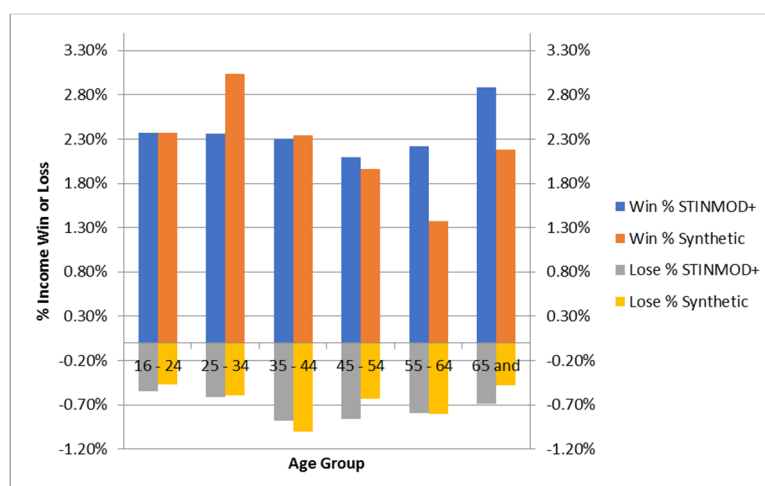

Figure 1. Results of a tax deduction for owner occupied properties, 2017-18, by Age of household head

4.2. Results by income decile

The results by household income decile are shown in Figure 2. It can be seen that STINMOD+ and the synthetic population gave similar results for all income deciles except the first, where STINMOD+ estimated double the loss compared to the synthetic population. Generally, lower income deciles gained, while the higher income deciles lost, which was expected.
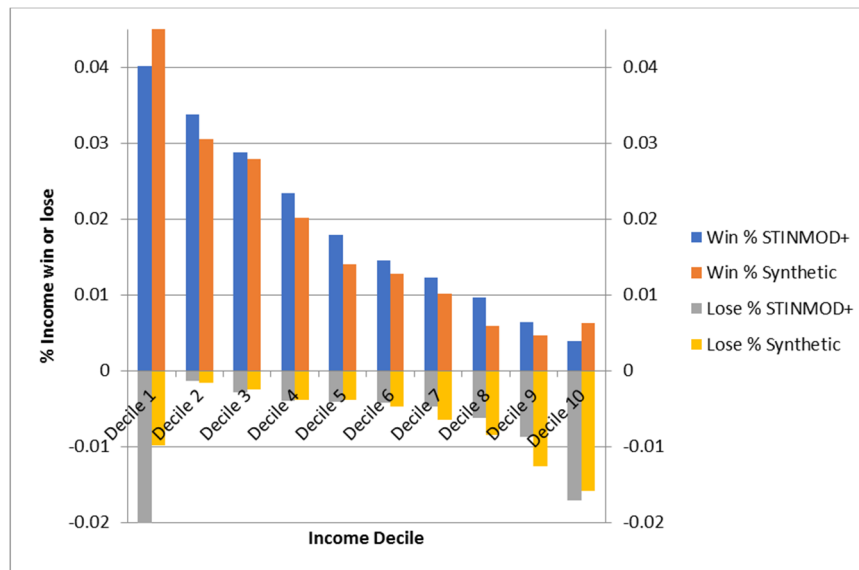
Figure 2. Results of a tax deduction for owner occupied properties, 2017-18, by income decile

4.3. Results by family type

The results by family type are shown in Figure 3. It can be seen that the difference between the STINMOD+ results and the synthetic results was very small, with the main difference for family types "Not Determined". This is a category that captures group households and households where the ABS has not been able to determine the family type. Generally one parent families with dependent children are the winners from this policy change, and Couple and multiple family households with dependent children are the losers, possibly due to the higher household incomes and therefore higher tax paid under the tax increase.
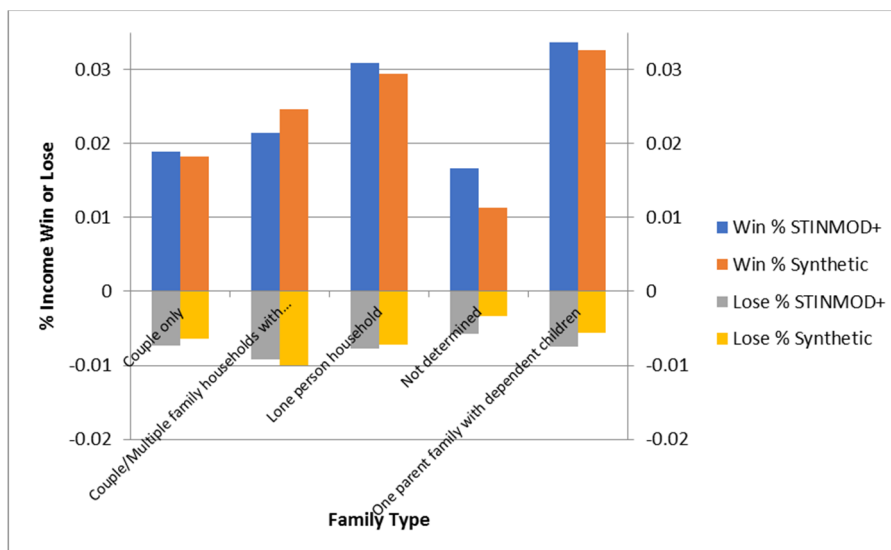

Figure 3. Results of a tax deduction for owner occupied properties, 2017-18, by family type

4.4. Results by SA2

The results for winners by SA2 are shown in Figure 4. In this map, we have used natural breaks to split the results into 5 categories. These categories represent the proportion of income won from the policy, so the highest winners (the dark blue areas) experience an increase of 2.5 – 3.4 per cent. It can be seen that areas along the East coast of Australia win from the policy.
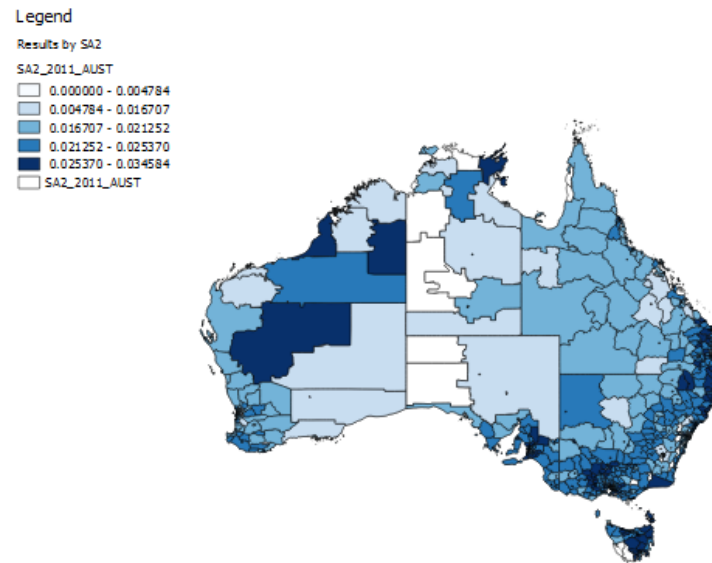
Figure 4. Results of a tax deduction for owner occupied properties, % Income gained by winners, 2017-18, by SA2

Figure 5 shows results for Sydney and Melbourne. In this map, lighter colours represent a greater loss (bad), and darker colours represent a lower loss (good). Most of the losers are in higher income capital cities, so only moderate amounts are lost across regional and rural areas in Australia. Three areas in remote WA show a high loss, and this may be due to higher incomes in these areas. Looking at South-West had lower losses probably due to lower incomes in these areas. Similar patterns were seen in Melbourne – areas of high incomes experienced greater losses, and areas of low incomes experienced lower losses. Essentially this is because the driver for the losses was income, while the driver for the gains was the mortgage payments.
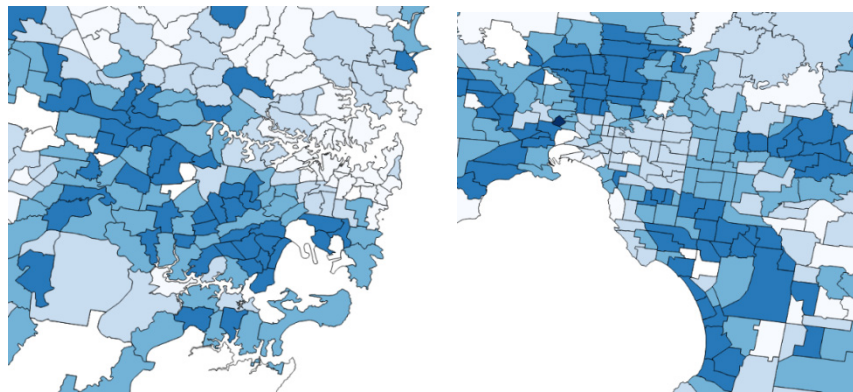

Figure 5. Results of a tax deduction for owner occupied properties, % Income lost by losers, Sydney (left) and Melbourne (right), 2017-18, by SA2

5. VISUALISING THE RESULTS

An important part of the results is visualizing the winners and losers by SA2. While the results in this paper show maps of winners and losers by SA2, because there is a synthetic population available for all SA2s in Australia, the results could also be visualized by winners and losers by SA2 by age group of the household head; by SA2 by family type; etc. This complex visualization can be achieved through the University of NSW Synt-Viz system (https://cityfutures.be.unsw.edu.au/cityviz/mortgage_taxation), and provides the researcher with the ability to map different cross-tabulations and visualize these results.

6. DISCUSSION AND CONCLUSIONS

This modelling has shown that a tax/transfer microsimulation model can be joined to a synthetic population using a direct matching of households in one model to households in the other model as long as the two models use the same base population. Other ways to link the models (for example, by identifying change in income for different groups of the population from the tax/transfer model and applying this change to each record of the same type from the spatial microsimulation model) can be used if the base population is not the same. The results show that the aggregated synthetic population results are similar to the aggregate tax/transfer microsimulation model results, with some differences for very low income households; and some age groups.

We expect that further experimentation with the two models will reduce these differences. The main limitation with this current work was that the tax/transfer model used two surveys, where the synthetic population used one. Further testing needs to be conducted with STINMOD+ using the same survey as the synthetic population, rather than two surveys; or the synthetic population needs to be created using the same two surveys as the tax/transfer model. Because the synthetic population is matched to 2011 Census data, all the survey income and expenditure data needs to be adjusted to 2011 prices. Looking at the results of the policy change, low income households benefited the most from the change; and high income households lost the most. This was mainly due to the tax increase. Areas in regional Australia gained the most, whereas areas in Australian cities (with higher incomes) lost the most due to the tax increase. Visualising these using the Synt-Viz system (Leao, et al. 2017) then provides researchers with the ability to look at results for different groups in the population.

These results were not the main aim of this paper. The main aim of this paper was to show how a tax/transfer model and spatial microsimulation model could be brought together by directly linking synthetic households in a synthetic population. This paper has succeeded in showing how this could work if the two models use the same base dataset, and other types of models that use unit record data (for example, an agent based model which might introduce interaction between agents, or interaction between agents and their environment) could also be linked into this system in the same way, as long as the agent based model used the same dataset. Further, aggregate economic models can also be integrated into the system using the second method discussed in this paper, as has been shown in Rao et al. (2015). The two models in theory could be brought together in the Geodesign system, used by the UNSW, as outlined in Tanton, Perez, et al. (2017), and visualisations could be developed in the Synt-Viz system. Visualisation is powerful front end for planners and policy-makers to interact with in better understanding space and place (Pettit, et al, 2012) The implementation of map based visualisation platform such as Synth-Viz and Geodesign would allow many other types of models to be integrated and viewed, providing the front-end model infrastructure to bring together different models to inform city planning.

REFERENCES

Burke, T., Stone, W., & Ralston, L. (2014). *Generational change in home purchase opportunity in Australia. AHURI Final Report No. 232, Australian Housing and Urban Research Institute Limited, Melbourne,.*

Department of Parliamentary Library. (2016). *Housing affordability in Australia*.

Harding, A., Vu, Q., Tanton, R., & Vidyattama, Y. (2009). Improving Work Incentives and Incomes for Parents: The National and Geographic Impact of Liberalising the Family Tax Benefit Income Test. *Economic Record*, *85*(s1), S48–S58. http://dx.doi.org/10.1111/j.1475-4932.2009.00588.x.

Harland, K. (2013). *Microsimulation model user guide (Flexible Modelling Framework)*.

Huynh, N., Barthelemy, J., Perez, P. (2016). A heuristic combinatorial optimisation approach to synthesising a population for agent-based modelling purposes. *Journal of Artificial Societies and Social Simulation* 19 (4), 11-1-11-21.

Huynh N., Perez P., Berryman M., Barthélemy J. (2015). Simulating Transport and Land Use Interdependencies for Strategic Urban Planning - An Agent-Based Modelling Approach Modelling Approach. *Systems*, 3 (4): 177-210.

Leao, S. Z., Huynh, N., Taylor, A., Pettit, C., & Perez, P. (2017). Evolution of a Synthetic Population and Its Daily Mobility Patterns Under Spatial Strategies for Urban Growth. In Planning Support Science for Smarter Urban Futures (pp. 399-417). Springer International Publishing.

Li and La (2017) Introduction to STINMOD+: A modern approach of tax and transfer system modelling in Australia, Technical Paper

Li, J., O'Donoghue, C., Loughrey, J., & Harding, A. (2014). Static Models. In C. O'Donoghue (Ed.), *Handbook of Microsimulation Modelling* (pp. 47–75). Emerald. doi:10.1108/S0573-855520140000293002.

Orcutt, G. H. (1957). A New Type of Socio-Economic System. *The Review of Economics and Statistics*, *39*(2), 116–123. doi:10.2307/1928528.

Pettit, C. Widjaja, I, Russo, P, Sinnott, R, Stimson, R, Tomko, M. (2012) Visualisation support for exploring urban space and place, XXII ISPRS Congress, Technical Commission IV 25 August – 01 September 2012, Melbourne, Australia Editor(s): M. Shortis, J . Shi, E. Guilbert, ISPRS Annals Vol 1-2, pp 153-158.

Rao, M., Tanton, R., & Vidyattama, Y. (2015). Modelling the economic, social and ecological links in the Murray Darling Basin : A conceptual framework. *Australasian Journal of Regional Studies*, *21*(1), 80–102.

Tanton, R. (2014). A review of Spatial Microsimulation Methods. *International Journal of Microsimulation*, *7*(1), 4–25.

Tanton, R., & Clarke, G. (2014). Spatial Models. In C. O'Donoghue (Ed.), *Handbook of Microsimulation Modelling* (pp. 367–383). Emerald. doi:10.1108/S0573-855520140000293011

Tanton, R., Huynh, N., Vidyattama, Y., & Barthelemy, J. (2017). Comparison of sample free and sample based synthetic populations. In *15th International Conference on Computers in Urban Planning and Urban Management*. Adelaide, 11 - 14 July, 2017.

Tanton, R., Perez, P., & Pettit, C. (2017). A framework for integrating collaborative city design with individual centred modelling. In *Agent Based Modelling of Urban Systems*. Sao Paulo, Brazil, May 8 - 9 2017.

Tanton, R., Vidyattama, Y., McNamara, J., Vu, Q., & Harding, A. (2009). Old, Single and Poor: Using Microsimulation and Microdata to Analyse Poverty and the Impact of Policy Change among Older Australians. *Economic Papers: A journal of applied economics and policy*, *28*(2), 102–120. http://dx.doi.org/10.1111/j.1759-3441.2009.00022.x

Tanton, R., Vidyattama, Y., Nepal, B., & McNamara, J. (2011). Small area estimation using a reweighting algorithm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(4), 931–951. doi:10.1111/j.1467-985X.2011.00690.x

Vidyattama, Y., Rao, M., & Tanton, R. (2014). Modelling the impact of declining Australian terms of trade on the spatial distribution of income. *International Journal of Microsimulation*, *7*, 100–126.

Williamson, P. (2007). CO Instruction Manual. Population Microdata Unit, Dept. of Geography, University of Liverpool.