MEASURING INEQUALITY FOR SMALL AREAS USING SPATIAL MICROSIMULATION: AN AUSTRALIAN CASE STUDY*

by Riyana Miranti, Rebecca Cassells, Yogi Vidyattama and Justine McNamara

In this paper, we have used spatial microsimulation techniques to calculate small area inequality in Australia using disposable income data which are not available at a small area level. Applying this technique, household synthetic data at small area level are created, drawing together data from the Australian Census and Survey. Using disposable income increases the strength of the results, as a more accurate measure of income distribution is able to be obtained. Small area inequality estimation enables the policy maker to pinpoint pockets of inequality and to link these with other small area characteristics. Further, a case study of New South Wales, the most populous state in Australia is analyzed and the results show that there are marked differences in what appears to be associated with variation in inequality between urban area (Sydney) and rural areas highlighting the complexity of income inequality at a small area level.

Keywords

Keywords: income inequality, spatial microsimulation, small area
JEL classification: D31, D63 and R12

* The authors are very grateful to the participants of the 2nd General Conference of the International Microsimulation Association, Ottawa, Canada, June 8 – 10, 2009 for useful suggestions and feedback on the preliminary version of this paper.
Correspondence : Riyana Miranti, Senior Research Fellow, National Centre for Social and Economic Modelling (NATSEM), University of Canberra. Email: riyana.miranti@natsem.canberra.edu.au.

I. *Introduction*

Measuring income inequality has long been of interest in applied social and economic research in Australia. Inequality can have social and political implications as Wilkinson (2006) argues this may create social and political conflict, violence, depressions and other issues. Knowing inequality is important because it helps policy makers to understand better the cause of inequality and may target policy program better.

In comparison to other OECD countries, the data in mid 2000s, shows that Australia sat at around the middle rank (ranked 19 out of 34 reported OECD countries from the highest to the lowest inequality) with a Gini coefficient of 0.315 which was close to the OECD inequality on average at 0.316. This was substantially higher than Denmark, which had the lowest inequality of 0.232, but much lower than Chile, which had the highest inequality of 0.503 (OECD 2013)[1]. While this provides us with a picture of where Australia falls internationally in terms of income inequality, much more can be said about the nature of inequality in Australia, and in this paper we focus on a sub-national analysis of Australia's income inequality.

Previous Australian studies mostly consider Australian income inequality at a national level, with only a few authors studying this phenomenon at a regional level. However, as there is increasing interest in studying regional diversity in inequality, as discussed by Athanasopoulous and Vahid (2003), policy makers are now interested in examining income inequality within and between regions (Gregory and Hunter 1995; Lyold et al. 2000; Chotikapanich et al. 2005). Just as the fruits of the recent economic boom are not spread evenly across all regions in Australia (Miranti et al. 2010; Saunders et al. 2008; Vu et al. 2008; Meagher and Wilson 2008), it is likely that the average Gini we see nationally is in fact

1 Gini coefficients are calculated for total population after taxes and transfers.

much higher (or lower) in some areas. There is some support for this notion in previous literature.

Internationally, there have been several studies that seek to measure inequality at broader geographic levels in order to analyze the regional disparities within a country. Trendle (2005) explains that the spatial variation of income inequality has been a debate within the field of regional science. Examples of this research includes Loikkanen et al. (2002) who examine inequality for Finland; Akita (2003) for China and Indonesia; Gray et al. (2003) for Canada; Balisacan and Fuwa (2003) for Philippines and Elbers et al. (2005) for Ecuador. However, there are few international studies that have sought to measure inequality at smaller geographic levels. Among these studies, Elbers, Lanjouw and Lanjouw (2003) and Tarozzi and Deaton (2009) have developed methods to calculate small area poverty and inequality in developing countries. Further, Ballas (2004) uses microsimulation to estimate the trends in poverty and inequality for two cities in England.

Some Australian research has attempted to measure income inequality at a small geographic area. Most of this research uses Statistical Division (SD) level data – a more aggregated geographical level than we use in this study. For example, Maxwell and Peter (1988) measure income inequality in 1976 and 1981 using Australian Census data and further examine the determinants of these inequality differences; McGillivray and Peter (1991) measure income inequality and examined its determinants in 1976, 1981 and 1986. O'Hagan (1999) examines income inequality and population movements in Victoria from 1981 to 1996 at a statistical division level. However, some current research focuses on more detailed regional disaggregation, although this has often been limited to selected states only. For example, Athanasopoulos and Vahid (2003) measure income inequality in selected and combined small areas (using statistical subdivisions) in New South Wales and Victoria (and some larger areas of other states) whilst, Trendle (2005) calculates income inequality and examines sources of regional income inequality for Local Government Areas (LGAs) in Queensland.

This research contributes to international research on small area inequality and expands the previous research for Australia with two improvements. First, this research explores inequality for small areas using equivalised household disposable income data calculated using spatial microsimulation techniques, whilst for example, Athanasopoulos and Vahid (2003) use gross income. As disposable income data is not available at a small area level, a spatial microsimulation model is used to calculate inequality. Using equivalised household disposable income increases the strength of the results, as a more accurate measure of living standards is able to be obtained as it measures resources available to households after paying income tax (Lyold et al. 2000) and a truer measure of inequality as argued in Harding (1997) there is some evidence that the income tax system has become more progressive and provides an offsetting force to growing inequality of gross income. Second, the unit of analysis used in this paper is the Statistical Local Area (SLA), which is a smaller geographical unit than any that has been used in previous Australian studies. Using a smaller spatial unit has several advantages, including the ability to pinpoint pockets of inequality, to link these with other small area characteristics, and to assist with effective policy and program targeting. An additional advantage of the methodology used in our analysis is that it gives a more precise measure of inequality, than traditional measures, due to the ability of spatial microsimulation to create the distribution disposable household income at small area level.

In this paper, inequality at a small area is estimated using conventional Gini coefficient methodology, and the Australian state of New South Wales (NSW) is used as a case study. NSW is the most populous state in Australia, housing Australia's most well known city – Sydney, and containing around 32.9 percent of Australia's total population in 2006. There are three main objectives of this research. Firstly, to provide valuable information about regional inequality at a small area level at a more disaggregated geographical level than what has been done previously. Secondly, to explore another use of spatial microsimulation and demonstrate its superiority in this area. And finally, to identify characteristics which previous research has found to be associated with inequality at a broader spatial level, and test if these relationships exist at a smaller level.

The remainder of this paper is organised as follows. The following section outlines the data and spatial microsimulation methodology used in this paper. Section III discusses how to measure inequality. Section IV describes the validation of our results, which are then discussed in Section V and finally Section VI provides a conclusion and policy implications.

II. *Data and Methodology*

i. Data

All data used in this study are originally sourced from the Australian Bureau of Statistics (ABS). All income data are sourced from the 2005-06 Survey of Income and Housing (SIH). For the spatial microsimulation analysis, the 2006 Census and the 2003-04 and 2005-06 ABS Surveys of Income and Housing are combined to maximize sample size. The spatial microsimulation methodology is discussed further in section II.2 Validation is conducted using national, state, and small area level data published by the ABS, using the 2006 Census and the latest available Confidentialised Unit Record Files (CURF) survey data (2005-06 SIH).

The 2003-04 SIH has a sample size of 11,361 households whilst the 2005-06 SIH has a sample size of 9,961 households. The sample used by the ABS for the SIH covers occupied private dwellings only. In contrast to the Census, the SIH has rich, detailed information about a range of socioeconomic variables, including disposable income. This detailed information on the SIH allows Gini coefficients to be generated using equivalised disposable income, in comparison to Census data which only provides income in ranges and available as *gross income* only. However, the SIH does not provide a detailed geographical disaggregation. Therefore, the SIH is suitable for analysing inequality at a larger geographical area such as national or state level, but not for small areas; while the Census is suitable for analysing many household characteristics at a small area level, but does not provide enough income data to create acceptable measures of inequality. Our spatial microsimulation techniques bring these two data sources together, using the Census to provide reliable small area benchmarks, which are used to reweight the SIH data.

The Gini coefficients are calculated at person level using household income, as we assume income sharing within households and, prior to conducting the analysis, negative household incomes are recoded to zero to follow the standard approach of the ABS (Li 2005). Li (2005) and Saunders et al. (2008) argue, some analysis has shown that the expenditure patterns of those households with zero and negative incomes are inconsistent with their reported low income.

Disposable household income is chosen since this is a better measure for income distribution analysis as it measures resources available to households after paying income tax (Lyold et al. 2000) and as argued in Harding (1997) there is some evidence that the income tax system has become more progressive and provides an offsetting force to growing inequality of gross income. Nevertheless, in the small area validation section, gross household income will be used for testing the spatial microsimulation results since the income in Census data is only available as gross household income. In common with other research, disposable household incomes are equivalised, so that rankings of income will then take into account the differences that household size and composition make to standards of living. Equivalence scales give ‘points’ to each adult and child in the household, and then the household’s disposable income is divided by the sum of these points so that incomes can be compared across different types of households. Here we use the modified OECD equivalence scale, which assigns the following values: 1.0 point for the first adult; 0.5 for each of the remaining adults and 0.3 for each dependent child in the household. It should be noted that for the purposes of calculating equivalised income, dependent children are defined as only those children aged less than 15 years, in common with current Australian practice.

The spatial unit used in this paper is the Statistical Local Area (SLA). The SLA is one type of standard spatial unit described in the Australian Standard Geographic Classification (ASGC) 2006 and is based on the boundaries of incorporated local government bodies where these exist (ABS 2007a). The 2006 Census data covered 1,426 SLAs in Australia which include 200 SLAs in NSW.

There are two main reasons why the SLA is used as the unit of analysis in this study. First, the SLA is the smallest unit in the ASGC where there are not substantial issues with confidentiality. Second, SLAs cover the whole of Australia (as opposed to other spatial unit such as Local Government Areas which do not cover areas with no local government) and cover contiguous areas (unlike some postcodes).

To examine the characteristics of these SLAs, data from the 2006 Census are used[2], including proportion of immigrants, indigenous, managers and professionals, female labor force participation, and proportion of people living in public housing.

ii. Spatial microsimulation methodology

Spatial microsimulation is essentially the calculation of a set of small area weights. By combining detail available on data-rich surveys with detail available on the geographically rich Census, we are able to create synthetic data that accurately estimate certain socio-

[2] At the time this paper was written, the latest 2011 Census has not been yet published.

economic phenomena that are closely related to the benchmarks which works as a predictor or determinant in the estimation to calculate these weights.

A set of data that is directly comparable between the survey and Census data is selected, and adjusted into appropriate cross-tabulations and groupings. These tables are known as "benchmark" tables, and currently comprise the variables shown in Table 1. As shown in Table 1, most of the benchmark tables are at household level, and only three of them are "person" level benchmarks.

Table 1 about here

Both the survey data and the Census are adjusted and manipulated in order to gain alignment for use in the reweighting process. Income values are uprated on both surveys using average weekly earnings in order to coincide with 2006 dollar values and mortgages and rents are also uprated using a factor derived from the Consumer Price Index. Extensive work has been undertaken for all benchmark components to ensure that they have the same definition and coverage on both the Census and the SIH (see Cassells et al. (2012; 2010) for more detail).

The reweighting process is carried out for the whole of Australia and followed the methodology described in Chin and Harding (2006) and Cassells et al. (2012; 2010). The procedure used is a SAS macro called GREGWT which uses an iterative constrained optimization technique to calculate weights that best represent all the Census benchmarks. The procedure is a generalised regression procedure outlined in Bell (2000). Because the reweighting process is an iterative process, there are areas where the procedure does not find a solution. If there is no solution found after 30 iterations, then the process has not converged. Those SLAs where the process does not converge are usually SLAs where the population is quite different to the sample population – for example, industrial estates or inner city areas.

However, for 29 SLAs within NSW, it is found that the GREGWT criterion for non-convergence is too strict: even after iterating 30 times and not converging, the estimates obtained from the weights were still reasonable when compared with the benchmarks. Therefore, a new criterion for reweighting accuracy, which uses the total absolute error (TAE) from all benchmarks is developed in order to maximize the number of SLAs for which we can produce valid data. With the latest criteria, if the absolute total error from all the benchmarks is greater than the population in that SLA, then the accuracy criteria has failed, and the SLA is dropped from any further analysis. Generally, the convergence criteria and the accuracy criteria provide the same results when an area has obviously not converged; but for marginal areas, the area may reach the maximum number of iterations but still provide a reasonable total absolute error. In the final results, TAE criteria are used rather than the GREGWT convergence criteria.

While the acceptance rate of SLAs is overall very high (especially when considered in population terms), we lose almost a third of the Northern Territory population in this reweighting process, however as this research concentrates on NSW, this does not affect the results.[3] Using the TAE criteria, only 2 SLAs are lost from NSW. This represents 0.34 per cent of the total NSW population (Table 2).

Table 2 about here

One SLA is further excluded from the sample where the estimated population size is less than 30 persons, as this population size is considered to be too small to produce reliable estimates. This additional criterion produces a total of 197 SLAs used to analyse inequality within NSW.

3. Only 0.79 per cent of the total Australian population in 2006 are lost in the reweighting process.

III. *Calculating Inequality*

There are various ways to measure inequality (see ABS (2006b) for a summary of measures of inequality including the Theil and Atkinson Index). This paper uses a Gini coefficient which measures disparity between each person in the population and every other person in the population through income. Gini coefficients are used to measure inequality for two reasons as follows (i) the Gini coefficient is the most commonly used summary measure (Athanasopoulous and Vahid (2003) and ABS (2006b)); and (ii) Gini coefficients are the only statistical measure of income distribution (at the national and state level) published by the ABS and thus allow us to validate our spatially microsimulated small area Gini coefficient estimates.

As discussed in Athanasopoulos and Vahid (2003, p. 414), the Gini coefficient satisfies the three basic criteria of acceptable inequality measures proposed by Sen and Foster (1997),

1) Scale independence: which states that if every individual's income in a society changes at the same proportion, the inequality measure in that society will not change;

2) Invariance to replication of population: which means if the population size is doubled by keeping the exact characteristics of the original population, the inequality measure will not change;

3) Compliance with the "Pigou-Dalton" principle of transfers: which means that inequality is expected to increase if there is an income transfer from a poorer person to a richer person and to decrease if the transfer is from a richer person to a poorer person (see Athanasopoulos and Vahid (2003) for further detail).

The Gini coefficient is often calculated using the following formula (ABS 2006b, pp. 6).

$$G = \left(\frac{1}{2n^2\mu} \right) \sum_{i,j}^{n} \left| y_i - y_j \right|$$

(1)

Where;

n = the number of people in the population

μ = the mean equivalised disposable household income of all people in the population

And y_i and y_j are the equivalised disposable household income of the ith and jth persons in the population.

Alternatively, the Gini coefficient can also be calculated by examining the Lorenz curve. The Lorenz curve is a curve with the horizontal axis showing the cumulative proportion of the persons in the population ranked according to their income and with the vertical axis showing the corresponding cumulative proportion of equivalised disposable household income.

As indicated in Figure 1, the Gini coefficient is a measure of the area between the Lorenz curve and the 45 degree line, calculated with this formula:

$$G = \left[\frac{A}{A+B} \right] \tag{2}$$

The Gini coefficient has a value between zero and one. A value of zero means perfect equality, a situation in which everyone in the population lives in a household with the same level of equivalised income. A value of one indicates perfect inequality, a situation where one person holds all the income. Smaller Gini coefficients indicate a more equal distribution of income.

Figure 1 about here

In this paper, as explained earlier, the household weights generated from the spatial microsimulation model are applied to calculate Gini coefficients at a small area. Therefore, the estimates of Gini coefficients are calculated using a weighted Gini formula as adapted in Harding and Greenwell (2001).

Applying household weights in each SLA to calculate Gini coefficients is challenging as there is no actual unit record data to calculate the Lorenz curve. Therefore, as highlighted in Harding and Greenwell (2001), the income distribution is determined by ranking people by their equivalised household income. Consequently, if a household has five people, their equivalised income will be counted five times, not once.

IV. *Validation*

As the estimates of the Gini coefficient are calculated using spatial microsimulation techniques, we undertake a set of validation procedures in order to check the accuracy of our synthetic estimates. Two procedures are used to validate the Gini coefficient estimates. These processes and results are described further below.

i. Small Area Validation

We used data directly from the 2006 ABS Census of Population and Housing as a point of comparison for our small area Gini coefficient estimates. The Census contains equivalised *gross* household income, and these data are only available in categories (in this case there are eight groups of income), not as a continuous variable, which limits our validation somewhat.

To calculate Gini coefficients using the Census data, the median value from each income range is calculated as the variable is categorical. Median income for each income range is calculated from the 2005-06 Survey of Income and Housing an applying the same income ranges used in the Census. All persons are allocated within each income range the same income value as we assume that there is an even distribution of income within each income range, and that the data are not skewed. For example, there are 1,617 persons within an SLA, that fall in the Census income category of $1-$149 per week. Thus, for the purpose of calculating the Gini coefficient, it is assumed that each of these 1,617 persons have the same equivalised gross household income of $73.10 which is the median equivalised income of this income range derived from the 2005-06 SIH.

It is noted that the Census validation data is limited by the categorical income groupings while the data derived from spatial microsimulation estimates is continuous and thus cover more groups of income. However, this is likely to result in Gini coefficient calculated from spatial microsimulation estimates being higher than the Gini coefficients calculated from the Census. This phenomenon is illustrated by Figure 2.

Figure 2 about here.

As shown in Figure 2, the "A_{sm}" area (the red-shaded area) which has six groups of income (please note that this is for an example only as the actual data has many more groups of income) calculated from spatial microsimulation is larger than the "A_{cs}" area (the orange-shade area) calculated from the Census data which has only three groups. As a consequence, applying the simple equation (2) to calculate the Gini coefficient, we will find that:

$$\left[\frac{A_{sm}}{A_{sm}+B_{sm}}\right] \rangle \left[\frac{A_{cs}}{A_{cs}+B_{cs}}\right] \Leftrightarrow G_{sm} \ \rangle \ G_{cs}\text{, since } A_{sm} \ \rangle \ A_{cs} \text{ and } A_{sm}+B_{sm}=A_{cs}+B_{cs} \quad (4)$$

where:

sm refers to spatial microsimulation and *cs* refers to Census.

Further, figure 3 shows the relationship between the Gini coefficients calculated using Census data (Y-axis) and those calculated using spatial microsimulation data (X-axis)

Figure 3 about here

Despite these differences, we find that our estimated Gini coefficients and Census Gini coefficients are closely matched for the majority of SLAs with a coefficient of determination value of 0.897. Thus we are confident that the spatial weights we have used are giving reasonable results for the small areas in NSW. Further, the Spearman rank correlation to measure the linear relationship between the two sets of ranked data is 0.958 for NSW, reflecting a high and positive linear relationship between these two ranked data.

ii. Aggregated Data Validation

The aggregated data validation is conducted using equivalised disposable household income data from the 2005-06 SIH. Income collected from the survey is argued to provide more accurate estimates of the distribution of income than income collected from the Census as interviewers are involved in collecting data directly from the survey respondents, whereas for the Census, the respondents complete the Census questionnaires without an interviewers' guidance (Maxwell and Peter 1988).

The Gini coefficients estimated at a SLA level are aggregated to the state and capital city and balance of state levels in order to compare these results directly with results available at this geographic level from the SIH.

Table 3 indicates that while the Gini coefficient estimates from spatial microsimulation tend to be higher than estimates from the 2005-06 SIH, the results are generally aligned[4]. The slightly higher estimates from SpatialMSM/09C are possibly due to benchmarking to Census data, which may have a greater amount of persons reporting lower incomes than the sample in the SIH, however overall the results are very promising. It is also found that the Gini coefficient estimates for the balance of each state have greater differences than those estimated for each capital city.

Table 3 about here

4 Results from Victoria (another populous Australian state) have also been included to show the aggregate patterns.

The aggregated data validation also shows that in general the capital cities have a higher Gini coefficient when compared to the balance of each state. This finding confirms previous research (Lloyd et al. 2000; Bray 2001). This may reflect that the capital cities in Australia are more heterogeneous in terms of income than the balance of state areas, as cities have a predominance of upper-middle income households together with very low income households (Lloyd et al. 2000).

V. *The Spatial Distribution of Inequality*

Figure 4 shows the natural breaking of inequality for New South Wales SLAs. The 197 areas are ranked and then divided into five categories according to where the greatest differences are in the data. Similarly figure 5 applies the same natural break categories for the whole of New South Wales to Sydney. The palest colour on the map represents areas that have the lowest income inequality category (the lowest Gini coefficient) while, in contrast, the darkest colour on the map represents areas with the highest income inequality (the highest Gini coefficient). The missing data on maps represents the excluded SLAs due to inadequate estimates and population sizes, as discussed earlier. The Gini coefficients in New South Wales vary between 0.262 and 0.369.

Figure 4 about here

Comparing figures 4 and 5, over 50 per cent of SLAs in the highest income inequality category (29 SLAs) lie within the capital city - Sydney (16 SLAs). SLAs with high income inequality are mostly clustered in Sydney, with some additional high inequality SLAs scattered throughout the state balance. In Sydney, these SLAs run in a horizontal corridor, from east to west, starting at the inner city suburbs of Waverley, Woollahra, Randwick, Ashfield and Strathfield, and flowing out along the western motorway (M4) and the major train line, towards the western suburbs of Auburn and Parramatta. For the balance of New South Wales, most SLAs with high inequality are concentrated in the remote far west and north western areas of New South Wales.

Figure 5 about here

i. Literature on Determinants of Inequality

Previous research has uncovered factors that are considered to be determinants of inequality in Australia (see for example Maxwell and Peter 1988; McGillivray and Peter 1991; Trendle 2005). These factors include the magnitude of particular groups that are often associated with low income, including immigrants, indigenous persons and public housing tenants; as well as those that are closely related with high income, such as those with 'high-end' occupations. We also examine the relationship between female labor force participation rates (FLPR) and inequality. Interestingly, previous literature has shown that the relationship of many of these variables to inequality, (particularly regional inequality) is often ambiguous.

In terms of immigrants, literature has discussed different groups of overseas born population in Australia. The first group is an older generation of immigrants such as Italian immigrants who arrived in Australia after WWII in the 1950s and Vietnamese who arrived in the mid-1970s after the Vietnam War, as discussed in Greig et al. (2003). These immigrants are characterized by limitations such as language barriers, low education levels, and racial discrimination, which in turn have resulted in a concentration of low paid occupations. This group of immigrants is more geographically dispersed and consequently their presence is expected to have a positive correlation with inequality. On the other hand, the relationship between another group of immigrants who are a younger generation that tends to be more concentrated in specific regions and earning similar levels of income. However, the literature discussed that relationship between this group of immigrants and inequality are ambiguous. It is expected that this group of people has a negative correlation with inequality (McGillivray and Peter 1991). Nevertheless, although it is noted that many of these younger generations of migrants are highly skilled, yet Greig et al. (2003) find that their qualification may not be readily recognized. In addition, these immigrants are likely to be concentrated in a few urban areas, and strongly attached to their community, and some experience limited labor market opportunities or what Greig et al (2003) argue as being limited by the "ethnic mobility trap" (Greig et al, 2003 pp.128), and consequently having a positive relationship with inequality.

The proportion of Indigenous persons in SLAs is argued to have a positive correlation with regional inequality as these populations tend to have lower levels of educational attainment, do not have as many opportunities to work in highly skilled occupations (Trendle 2005), and are more likely to be unemployed and dependent on government benefits.

Previous research has found a positive relationship between female labor force participation and inequality with two different reasons. However, these differences are likely to be due to the differing types of income used to measure inequality. Trendle (2005) finds that female labor force participation has a positive correlation with inequality (measured using personal income) as this group of workers may have a higher probability of experiencing a career break/disruption or to work part time in order to care for children or families and earn lower incomes. However, McGillivray and Peter (1991) argue if those females who enter a labor market in a region are married, the number of families with double incomes in that particular region will also increase, which is expected to lead to greater inequality (measured using family income).

Much of the previous literature discusses the relationship between a country's income distribution and the stage of economic development (Kuznet 1955; Williamson 1965 for international literature and Amos 1986; Maxwell and Peter 1988 and McGillivray and Peter (1991) for Australia). The stage of development is usually represented by 'real' average income which is adjusted to reflect different costs of living across areas in Australia (Maxwell and Peter 1988). Since it is difficult to find a comprehensive cost of living index, Maxwell and Peter (1988) suggest using the proportion of the population with a post school qualification to represent the level of development. A high proportion of this variable may represent a higher level of regional development (Maxwell and Peter 1988). The proportion of the population with post school qualifications is expected to have a negative association with inequality (Maxwell and Peter 1988; Trendle 2005). Nevertheless, the relationship between inequality and the proportion of people with post school qualifications may also be ambiguous as Amos (1986) as cited in McGillivray and Peter (1991, pp. 137) argues, there is an "augmented inverted U" relationship between income inequality and the level of development, whilst McGillivray and Peter (1991, pp. 140) argue "short-run oscillations" between income inequality and the level of development which proposes a possibility of positive association between income inequality and the level of development. Glaeser, Resseger and Tobio (2008) find that an increase in the share of college graduates increases inequality (using the Gini coefficient as a proxy) in urban areas. In this paper, we use the proportion of people with at least a bachelor degree to represent post school qualifications.

The proportion of the population who are employed as either a manager or professional represents a highly skilled workforce, and these persons generally have higher incomes than employees in other occupations (Lyold, 2000). People who are employed as managers and professionals are also most probably those people with a post school qualification. Glaeser, Resseger and Tobio (2008) argue that the inequality of human capital within a region is related to returns to skill, and that those with a post school qualification will be concentrated in places where the return to human capital is higher (mostly in urban areas). Thus, it is also likely these persons with highly skilled occupations are concentrated together, and follow a similar pattern for those people who have a post school qualification, thus areas with a high proportion of people employed in professional and managerial occupations may also have high inequality.

We are unable to find any literature detailing the relationship between inequality and public housing, however in the Australian context, it is likely that persons living in public housing also have very low incomes. The presence of public housing within small areas is likely to increase income inequality, as it increases the number of persons in the lower end of the household income distribution, however this will also depend upon where the public housing is primarily located.

ii. Relationship between Regional Characteristics and Inequality

Table 4 shows the average proportion of persons in each Gini coefficient group by selected characteristics for all of NSW. It can be seen that SLAs in the highest inequality group are characterized by, on average, high proportions of immigrants, Indigenous persons, people working as managers and professionals, female labor force participation, and public housing tenancy (in comparison to other Gini coefficient groups). However, it is interesting that the lowest inequality group in NSW also has a high proportion of immigrants, and high female labor force participation.

Table 4 about here

Table 5 shows some clear differences in characteristics between Sydney and the balance of NSW (rural areas of NSW), and enables the patterns in Table 4 to be further fleshed out. The average proportion of immigrants has an opposite relationship for Sydney and rural NSW, with the highest average population of immigrants in the balance of NSW in the lowest inequality group, yet for Sydney, the highest average proportion of immigrants is in the areas of the city with the highest levels of income inequality.

Table 5 about here

The effect of average Indigenous populations for each area also differs, with an increasingly larger average proportion of Indigenous persons as inequality increases. For the highest inequality group in the balance of NSW, around 15.2 per cent of the population is Indigenous, whereas, for the lowest grouping, only around 3.7 per cent of the population is Indigenous. This is a relationship that is supported in the literature, where the proportion of Indigenous persons in SLAs is argued to have a positive correlation with regional inequality as these populations tend to have lower levels of educational attainment, as Trendle (2005) said that they do not have as many opportunities to work in highly skilled occupations and are more likely to be unemployed and dependent on government benefits. For Sydney, overall, there are much lower average proportions of Indigenous persons throughout the SLAs, ranging from only 0.7 to 1.7 per cent, and the pattern of Indigenous persons for each inequality grouping appears to be the reverse of that of the balance of NSW, with the highest proportion of Indigenous persons in the lowest inequality grouping. This may be because they live in the inner-city suburbs of Sydney where lots of rich people live, therefore high proportion of Indigeneous does not make any difference to the level of inequality.

The greater presence of persons working at higher occupational levels (managers and professionals), has a positive relationship with inequality levels - increasing as inequality does for both the balance of NSW and Sydney. This is supported by research from Glaeser, Resseger and Tobio (2008) who argue that the inequality of human capital within a region is related to returns to skill, and that those with a post-school qualification will be concentrated in places where the return to human capital is higher (mostly in urban areas). Thus, it is also likely these persons in highly skilled occupations are concentrated together, and follow a similar pattern for those who have a post school qualification, thus areas with a high proportion of people employed in professional and managerial occupations are likely to have high levels of inequality.

Whilst previous research has found a positive relationship between female labor force participation and inequality, we find that the average female labor force participation has a U shaped pattern for both Sydney and the balance of NSW, where the highest average participation rates are in both the lowest and highest income inequality groupings. The high average female labor force participation in both the lowest and highest inequality groupings is likely to represent two different female populations. For the highest inequality area in Sydney, it is likely that the group of women living here are mostly single or partnered professionals, working full-time without children. For the lowest income inequality areas, yet high female labor force participation, the group of women here are likely to be women with children, working part-time and living in the outer suburbs.

The presence of public housing within small areas is likely to increase income inequality, as it increases the number of persons in the lower end of the household income distribution; however this will also depend upon where the public housing is primarily located. Thus from Table 5, it is found for Sydney, public housing fluctuates between groups and there is no clear pattern. This could be due to public housing being located primarily in areas that have not too dissimilar private housing income distributions. However, for rural NSW, the highest inequality grouping also has the highest average proportion of persons in public housing – at 5.2 per cent.

VI. *Conclusion and Policy Implications*

This research has demonstrated that spatial microsimulation is an effective analytical tool that can be used to enhance estimates of inequality at small geographic levels. The validation process shows that our results are robust, and are closely aligned with small area estimates from the 2006 Census, using equivalised gross household income, and aggregated data obtained from the spatial microsimulation process, compared with direct estimates calculated from the 2005-06 Survey of Income and Housing.

Although there are slight differences in terms of magnitude, the weights give reasonable results for the vast majority of small areas in NSW with the broad regional rankings being very similar across both the Census and the synthetic estimates. The aggregation to state average shows that whilst the Australian Gini coefficient is 0.308, it is higher for NSW (0.322).

Our research illustrates that there are clear groupings of SLAs with high income inequality both in Sydney and rural NSW, providing useful information for policy makers and program targeting. While a comprehensive multivariate analysis of the determinants of inequality is beyond the scope of this paper, further analysis of the inequality groupings are carried out to examine various characteristics of small areas which previous research have found to be correlated with inequality.

Our findings show that, as expected, the highest inequality small areas in NSW are characterized by a mixture of people at the higher end of the income distribution (such as those working as managers and professionals) and the bottom of the income distribution (such as the proportion of the population living in public housing).

The findings of this research show that inequality does differ considerably when drilling down to smaller spatial areas. These findings are important, as they can aid in making regional policy aimed to reduce inequality more effective. By knowing which small areas are more unequal in regards to income, together with the characteristics of these areas, policy makers and service providers are able to better understand the intricacies of inequality and identify possible drivers, which can aid in more efficient targeting of programs and policy.

References

ABS (Australian Bureau of Statistics). (2006a), *Income and Housing Survey (2005-06)*, 2nd Edition, Basic Confidentialised Unit Record Files.

ABS (Australian Bureau of Statistics). (2006b), *Household Expenditure Survey and Survey of Income and Housing: User Guide*, 2003-04, cat no. 6503.0.

ABS (Australian Bureau of Statistics). (2007a), *Australian Standard Geographical Classification* (ASGC), cat. no. 1216.0.

ABS (Australian Bureau of Statistics). (2007b), *Household Income and Income Distribution*, cat no. 6523.0.

ABS (Australian Bureau of Statistics). (2008), *Australian Social Trends*, cat no. 4102.0.

Akita, T. (2003), 'Decomposing regional income inequality in China and Indonesia using two-stage nested Theil decomposition method,' *The Annals of Regional Science*, 37, 55–77.

Amos, O.M., Jr. (1986), 'Substate and SMSA Personal Income Inequality and Regional Development,' *The Review of Regional Studies*, 16, 23-30.

Athanasopoulos, G. and Vahid, F. (2003), 'Statistical Inference and Changes in Income Inequality in Australia,' *The Economic Record*, 79, 412-24.

Ballas, D. (2004), 'Simulating Trends in Poverty and Income Inequality on the Basis of 1991 and 2001 Census Data: a Tale of Two Cities,' *Area*, 36, 146-63.

Balisacan, A. and Fuwa, N. (2006), 'Changes in spatial income inequality in Philippines: An exploratory analysis,' in Kanbur, R., Venables, A.J. and Wan, G.H. (eds.), *Spatial Disparities in Human Development: Perspective from Asia*, University Press, United Nations, Tokyo; 207-32.

Bell, P. (2000), *GREGWT and TABLE Macros - Users Guide*, Unpublished, Australian Bureau of Statistics, Canberra.

Bray, J.R. (2001), 'Social Indicators for Regional Australia,' *Department of Families, Housing, Community Services and Indigenous Affairs Policy Research Paper* No. 8.

Cassells, R., Miranti, R., and Harding, A. (2012) 'Building a static spatial microsimulation model: data preparation', in Tanton, R. and Edwards, K. (eds.), *Spatial Microsimulation: A Reference Guide for Users*, Springer, UK; 9-16.

Cassells, R., Harding, A., Tanton, R., Miranti, R. and McNamara, J. (2010), 'Spatial Microsimulation: Preparation of Sample Survey and Census Data for SpatialMSM/08 and SpatialMSM/09,' *NATSEM Technical Paper* No. 36.

Chin, S.F. and Harding, A. (2006), 'Regional Dimensions: Creating Synthetic Small-area Micro Data and Spatial Microsimulation Models,' *NATSEM Technical Paper* No. 33.

Chotikapanich, D., Flatau, P., Adam, M. and Lloyd, R. (2005), 'Trends in Income Inequality and Poverty in Australia in the 1990s: Regional Perspectives,' paper presented at the Australian Social Policy Conference Sydney, 20-22 July.

Elbe's, C., Lanjow, J., and Lanjow, P. (2003), 'Micro-level estimation of poverty and inequality,' *Econometrica*,71, 355-64.

Elbers, C., Lanjow, P., Mistiaen, J., Ozler, B. and Simler, K. (2005), 'Are neighbours equal? Estimating local inequality in three developing countries,' in Kanbur, R. and Venables, A.J. (eds.), *Spatial Inequality and Development*, Oxford University Press, Oxford; 37-60.

Glaeser, E.L., Resseger, M. and Tobio, K. (2008), 'Urban Inequality,' *NBER Working Paper* No. 14419.

Gray, D., Mills, J.A. and Zandvakili, S. (2003), 'New Perspective on Canadian National and Provincial Income Inequality,' paper presented at the UNU-WIDER Conference on Inequality, Poverty and Human Well-being, Helsinki, 30-31 May.

Greig, A., Lewins, F. and White, K. (2003), *Inequality in Australia*, Cambridge University Press, Cambridge,.
Gregory, H. and Hunter, B. (1995), 'The Macro Economy and the Growth of Ghettos and Urban Poverty in Australia,' *Centre for Policy Research Discussion Paper*, No. 325.
Harding, A. (1997),'The Suffering Middle: Trends in Income Inequality in Australia,' *NATSEM Discussion Paper*, No. 21.
Harding, A. and Greenwell. H. (2001), 'Trends in Income and Expenditure Inequality in the 1980s and 1990s,' paper presented at the 30th Annual Conference of Economists, 24 September.
Kuznet, S. (1955), 'Economic Growth and Income Inequality,' *American Economic Review*, 45, 1-28.
Li, Y. (2005), 'Impact of Demographic and Economic Changes on Measured Income Inequality,' paper presented at the Australian Social Policy Conference 2005, Sydney.
Lloyd, R., Harding, A. and Hellwig, O. (2000), 'Regional Divide? A Study of Incomes in Regional Australia,' *Australasian Journal of Regional Studies*, 6, 271-92.
Loikkanen, H.A., Riihela, M. and Sullstrom, R. (2002), 'Regional income convergence and inequality in boom and bust: results from micro data in Finland 1971-1998, and especially during the 1990s,' paper presented at the meeting of the ESRC Urban and Regional Economics Seminar Group, University of Central Lancashire, UK 1-3 July.
Maxwell, P. and Peter, M. (1988), 'Income Inequality in Small Regions: A Study of Australian Statistical Division,' *The Review of Regional Studies*, 18, 19-27.
McGillivray, M. and Peter, M. (1991), 'Regional Income Inequality in A Developed Nation: A Cross-Sectional Study of Australian Sub-State Regions', *The Review of Regional Studies*, 21, 137-51.
Meagher, G. and Wilson, S.(2008), 'Richer, but More Unequal: Perceptions of Inequality in Australia 1987-2005,' *Journal of Australian Political Economy*, 61, 220-43.
Miranti, R., Harding, A., McNamara, J., Ngu, Q.V. and Tanton, R. (2010), 'Children with Jobless Parents: National and Small Area Trends for Australia in the Past Decade', *Australian Journal of Labour Economics*, 13, 27-48.
OECD. (2013), http://www.oecd-ilibrary.org/statistics; data extracted on 29 Apr 2013.
O'Hagan, R. (1999), 'Income Disparities and Population Movements in Victoria,' *The Australasian Journal of Regional Studies*, 5, 87-100.
Saunders, P., Hill, T. and Bradbury, B. (2008), 'Poverty in Australia: Sensitivity Analysis and Recent Trends', *Social Policy Research Centre Report*, No. 4.
Sen, A and Foster, J.(1997), *On Economic Inequality*, Clarendon Press, Oxford.
Trendle, B.(2005), 'Sources of Regional Income Inequality: An Examination of Small Regions in Queensland,' *Review of Urban and Regional Development Studies*, 17, 35-50.
Tarozzi, A. and Deaton, A. (2009), 'Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas, *The Review of Economics and Statistics*, 91, 773-92.
Vu, Q.N., Harding A., Tanton, R., Nepal, B. and Vidyattama, Y.. (2008), 'Advance Australia Fair? ' *AMP NATSEM Income and Wealth Report*, No. 20.
Williamson, J. (1965), 'Regional Inequality and the Process of National Development,' *Economic Development and Cultural Change*, 4, 3-47.
Wilkinson, R. (2006), *The Impact of Inequality: How to Make Sick Societies Healthier*, New Press, New York, London.

Table 1

Benchmark tables used for SpatialMSM/09C

No	Benchmark Table	Level
1	All household type	Household
2	Age by sex by labour force status	Person
3	Tenure by weekly household rent	Household
4	Tenure by household type	Household
5	Tenure by weekly household income	Household
6	Persons in non-private dwellings	Person
7	Monthly household mortgage by weekly household income	Household
8	Dwelling structure by household family composition	Household
9	Number of chidren aged under 15 usually resident in household	Household
10	Number of adults usually resident in household	Household
11	Weekly household rent by weekly household income	Household
12	Gross equivalised weekly household income by age	Person

Source: ABS Census Population and Housing 2006

Table 2

Number and characteristics of failed SLAs

State/Territory	Total SLAs	Failed SLAs	Proportion of failed SLAs	Proportion of persons living in failed SLAs out of all persons within state/territory
New South Wales	200	2	1.0%	0.34%
Victoria	210	7	3.3%	0.52%
Queensland	479	45	9.4%	0.75%
South Australia	128	7	5.5%	0.32%
Western Australia	156	17	10.9%	0.87%
Tasmania	44	2	4.5%	0.15%
Northern Territory	96	53	55.2%	28.37%
Australian Capital Territory	109	16	14.7%	0.61%
AUSTRALIA	1422	149	10.5%	0.79%

Source: SpatialMSM/09C applied to SIH2003-04 and SIH2005-06, ABS Census Population and Housing 2006

Table 3

Comparison of Gini coefficient estimates from the 2005-06 Survey of Income and Housing and SpatialMSM/09C

State	Capital city/ Balance of state	SpatialMSM/09C	2005-06 SIH	
New South Wales	All	0.322 (+)	0.317	
	Sydney	0.324(+)	0.321	*
	Balance of state	0.300(+)	0.287	*
Victoria	All	0.306	0.306	
	Melbourne	0.308(-)	0.309	*
	Balance of state	0.290(+)	0.274	*
Australia		0.308(+)	0.307	

Note: + (-) indicates where the estimates from spatial microsimulation are higher or lower than the estimates directly from 2005-06 SIH; *indicates that the coefficients have been calculated by authors. The Gini data at the capital city and balance of state level are not available from the ABS publication.
Source: ABS (2007b; 2008) and SpatialMSM/09C applied to 2003-04 and 2005-06 SIH, ABS 2006 Census of Population and Housing.

Table 4

Average proportion of persons in each Gini coefficient group by selected characteristics, all NSW, 2006

Gini coefficient	Immigrants	Indigenous	Managers and professionals	Female LFPR	Public housing
	%	%	%	%	%
0.262 – 0.280 (22)	16.37	2.61	28.56	56.25	3.42
0.281 – 0.292 (51)	12.21	3.27	30.59	51.28	3.94
0.293 – 0.304 (44)	12.73	3.06	33.97	51.12	3.54
0.305 – 0.321 (51)	15.53	4.62	39.70	52.04	3.33
0.322 - 0.369 (29)	23.79	7.19	46.17	54.59	4.51

Source: ABS Census Population and Housing 2006

Table 5

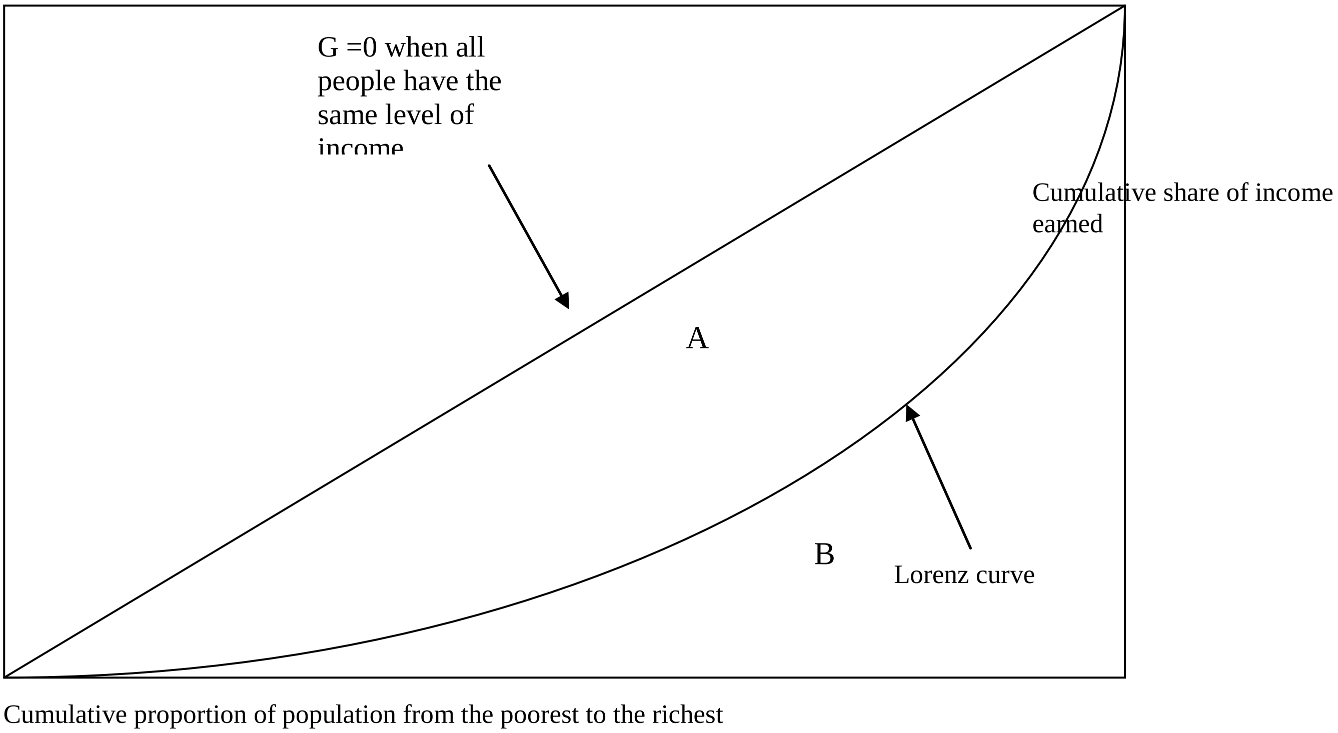
Average proportion of persons in each Gini coefficient group by selected characteristics, balance of NSW and Sydney, 2006

Gini coefficient	Immigrants		Indigenous		Managers and professionals		Female LFPR		Public housing	
	NSW - balance	Sydney	NSW - balance	Sydney	NSW - balance	Sydney	NSW - balance	Sydney	NSW - balance	Sydney
0.262 – 0.280	9.05	22.46	3.66	1.73	28.49	28.61	53.00	58.95	3.46	3.39
0.281 – 0.292	8.68	25.22	3.74	1.54	30.86	29.60	50.65	53.58	3.34	6.14
0.293 – 0.304	8.23	33.00	3.61	0.59	33.64	35.48	50.56	53.67	3.54	3.58
0.305 – 0.321	5.99	36.40	6.50	0.52	38.86	41.52	51.65	52.89	2.96	4.13
0.322 - 0.369	5.90	38.34	15.17	0.71	44.66	47.39	54.44	54.70	5.17	3.98

Source: ABS Census Population and Housing 2006

Figure 1

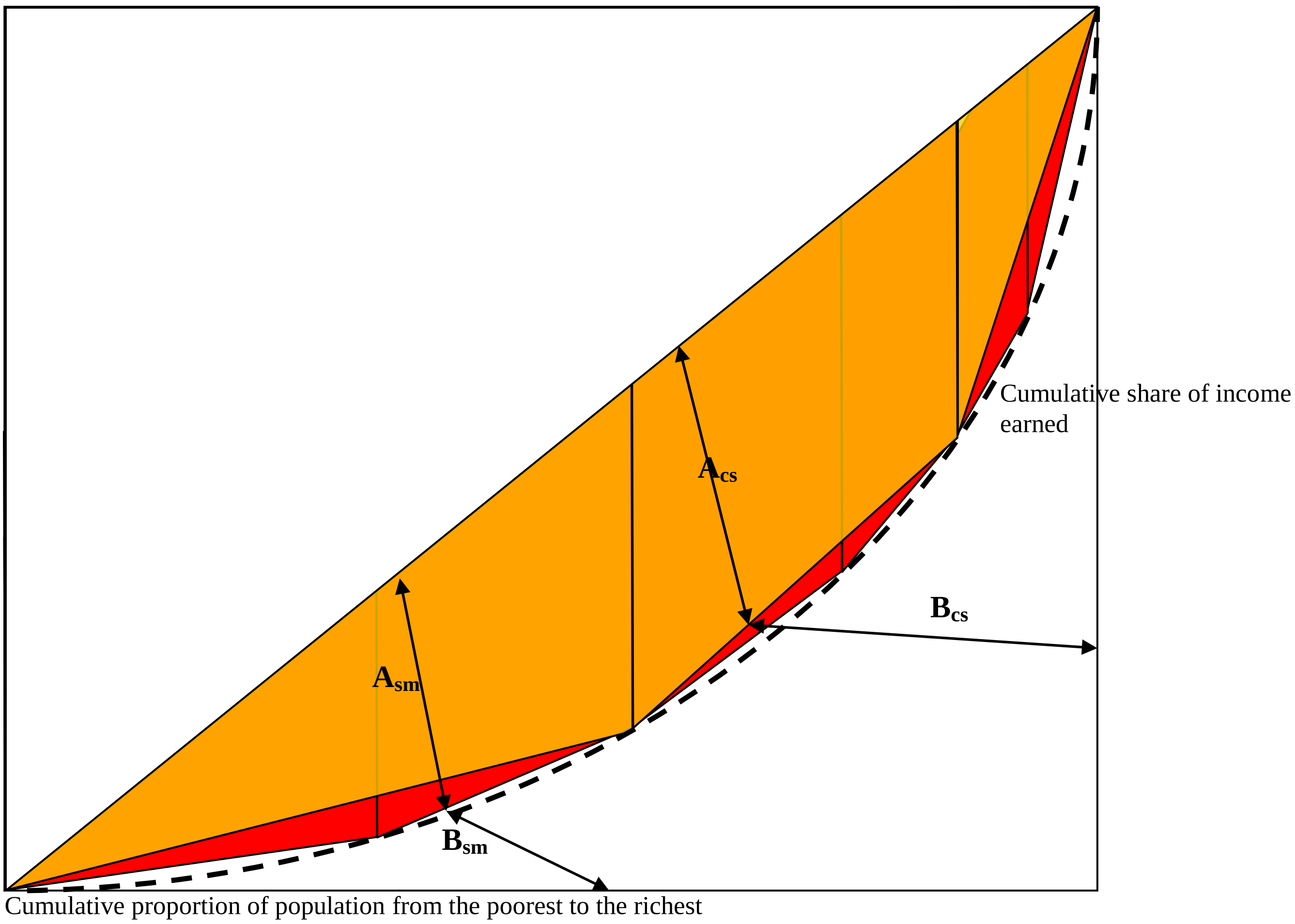
Lorenz Curve



Source: modified from Athanasopoulos and Vahid (2003)

Figure 2

Comparison of Gini Coefficients estimated from spatial microsimulation and Census data

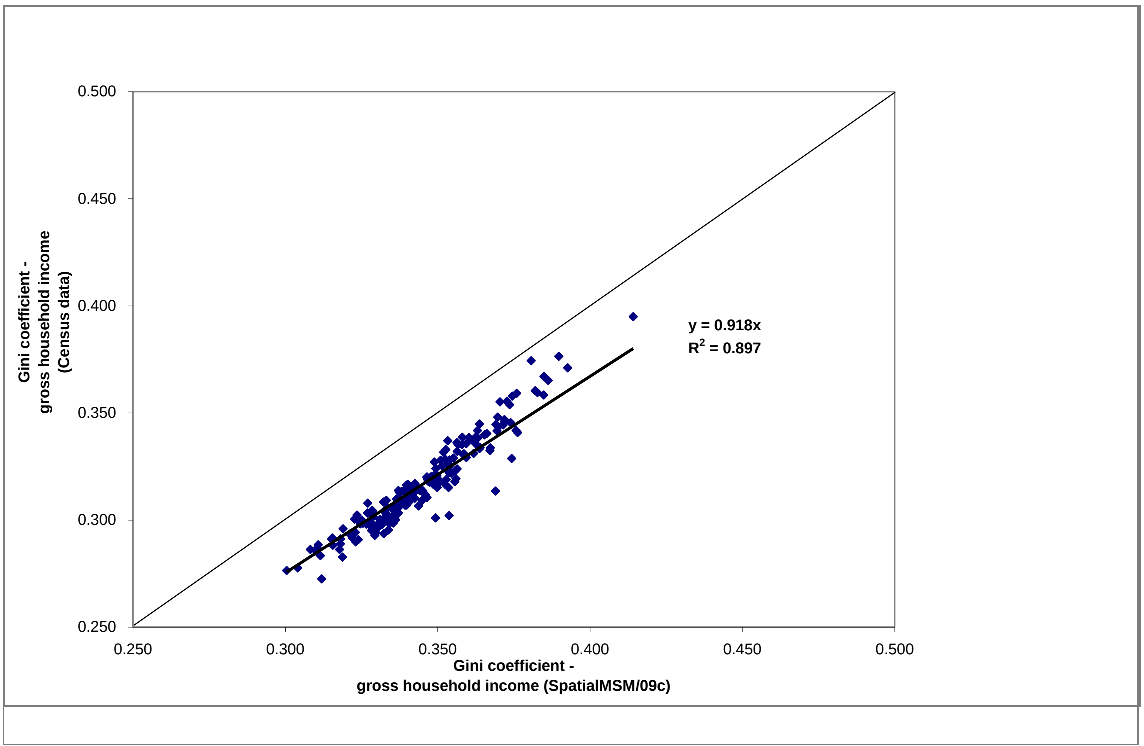


Note: *sm* refers to spatial microsimulation and *cs* refers to Census. $A_{sm} + B_{sm} = A_{cs} + B_{cs}$

Source: Authors' calculation

Figure 3

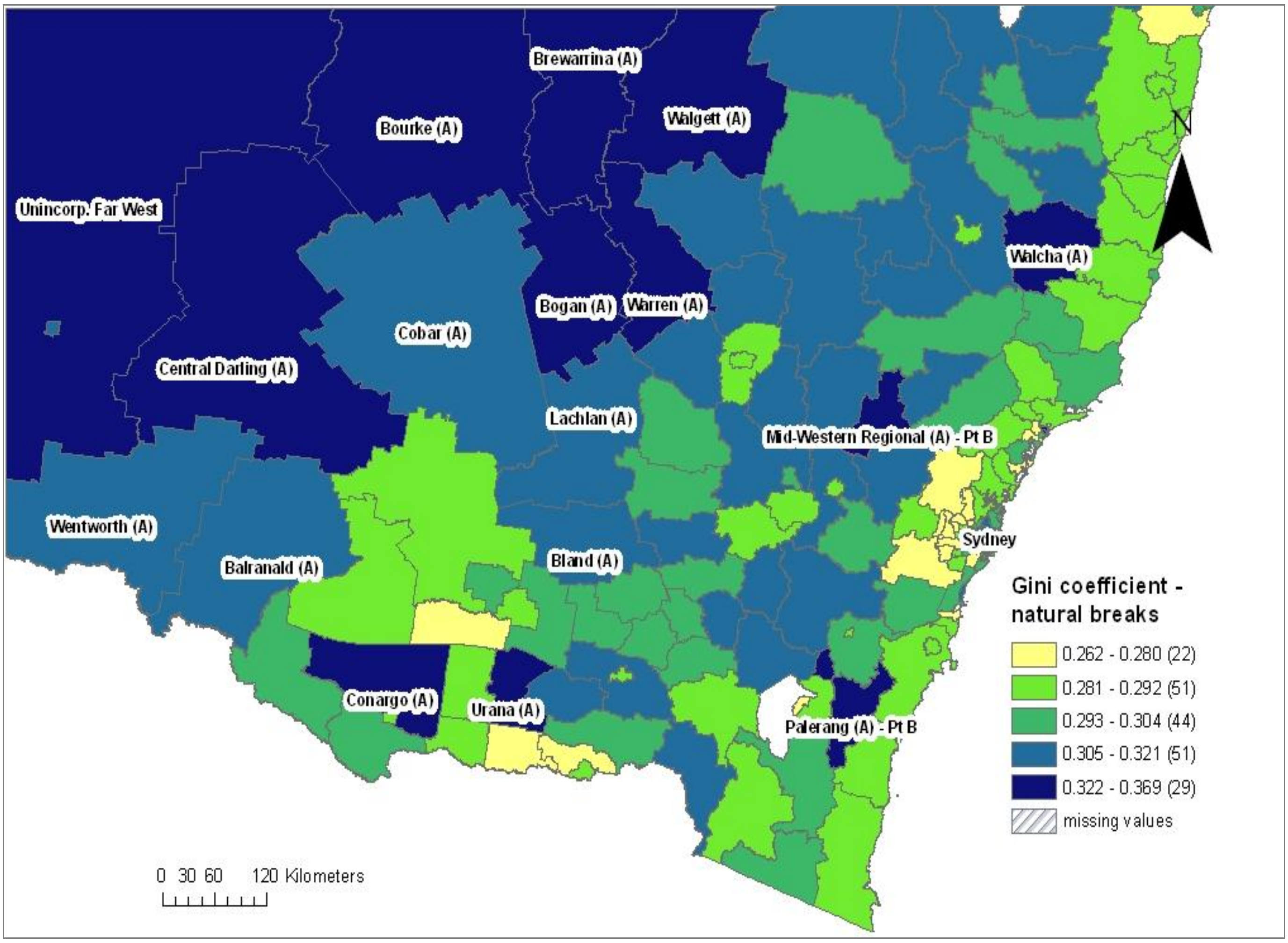
Comparison between Census Gini coefficients and spatial microsimulation estimates for persons, New South Wales, 2006



Source: SpatialMSM/09C applied to 2003-04 and 2005-06 SIH, ABS Census Population and Housing 2006

Figure 4

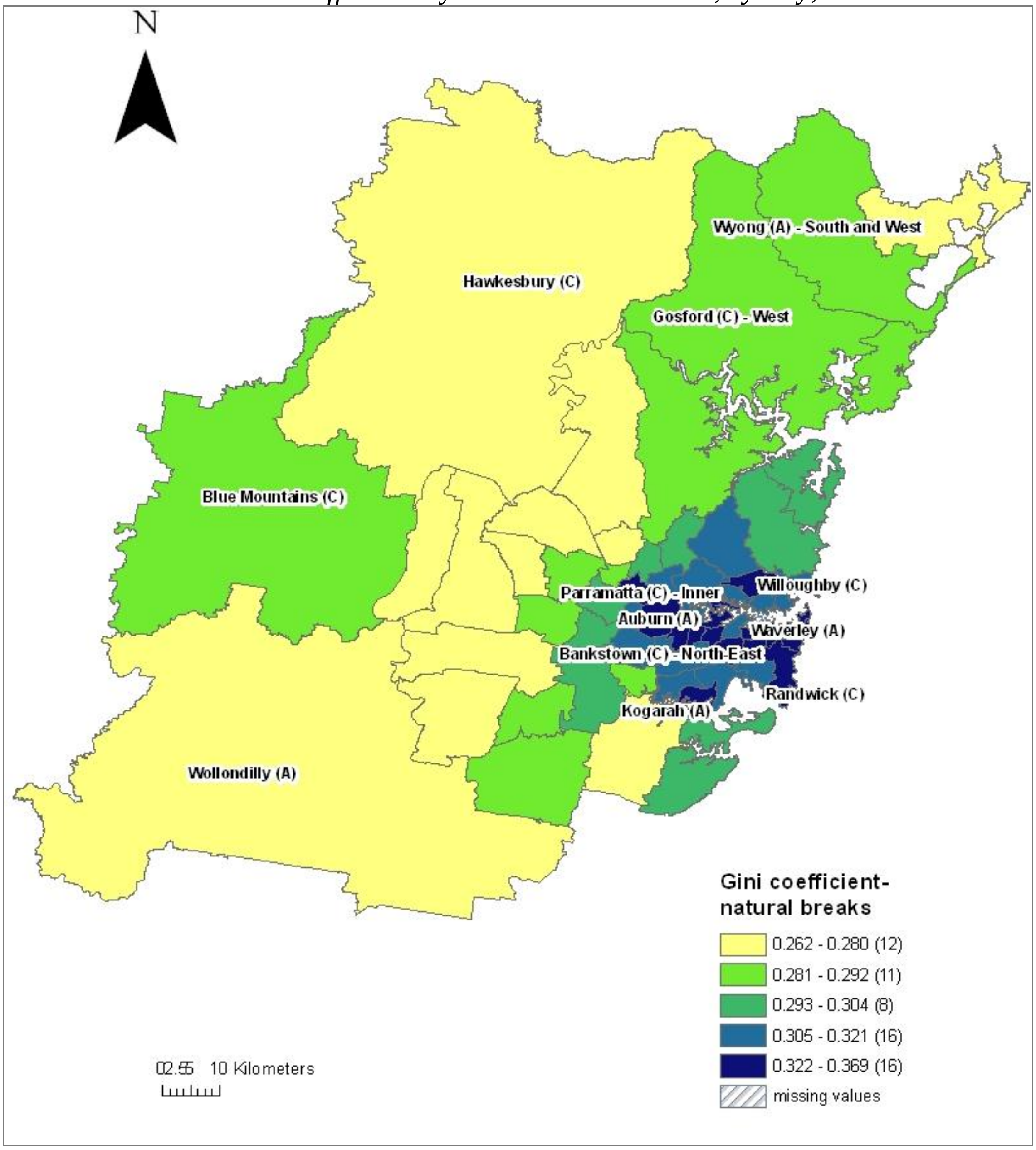
Gini Coefficients by Statistical Local Area, New South Wales, 2006



Source: SpatialMSM/09C applied to 2003-04 and 2005-06 SIH, ABS Census Population and Housing 2006

Figure 5

Gini Coefficients by Statistical Local Area, Sydney, 2006



Source: SpatialMSM/09C applied to 2003-04 and 2005-06 SIH, ABS Census Population and Housing 2006