Stability of Incomes for small geographies

REPORT FOR THE NATIONAL SCHOOL RESOUCING BOARD

MARCH 2018

By Robert Tanton and Dominic Peel

Note-this version of the report does not contain appendices. To view the full report, please access the following link

About NATSEM/IGPA

The National Centre for Social and Economic Modelling (NATSEM) was established on 1 January 1993, and supports its activities through research grants, commissioned research and longer term contracts for policy analysis and model development and maintenance.

In January 2014, the Institute for Governance and Policy Analysis (IGPA) at the University of Canberra was established to harness the research strengths of NATSEM and the ANZSOG Institute for Governance (ANZSIG). The aim of this Institute is to create and sustain an international class research institution for the study and practice of governance and public policy. The Institute has a strong social mission committed to the production of leading edge research and research driven education programs with genuine public value and, by implication, policy impact. The integration of ANZSIG and NATSEM has created exciting opportunities for the development of cutting edge research in public policy analysis through combining expertise in qualitative and quantitative methods, micro-simulation and policy modelling and evaluation.

NATSEM is one of three research centres within IGPA. NATSEM aims to be a key contributor to social and economic policy debate and analysis by undertaking independent and impartial research of the highest quality, including supplying valued commissioned research services. NATSEM is one of Australia's leading economic and social policy research centres and is regarded as one of the world's foremost centres of excellence for micro-data analysis, microsimulation modelling and policy evaluation. In keeping with IGPA's core mission, NATSEM's research activities aim to have significant policy impact and lead to social and economic change.

IGPA Director: Professor Mark Evans

NATSEM Directors: Professor Robert Tanton and Professor Laurie Brown

TABLE OF CONTENTS

Author Note

Authors of this report are:

Professor Robert Tanton, University of Canberra

Dominic Peel, University of Canberra

Acknowledgement

This study is undertaken by the National Centre for Social and Economic Modelling (NATSEM), the Institute for Governance and Policy Analysis (IGPA), at the University of Canberra, and was commissioned by the National School Resourcing Board.

We would like to thank Nic Biddle and Andrew Wade for comments on earlier versions of this report.

Acronyms

ABS	Australian Bureau of Statistics
ATO	Australian Taxation Office
PIT	Personal Income Tax
RMSE	Root Mean Squared Error
SA1	Statistical Areas Level 1
SA2	Statistical Areas Level 2
SEIFA	Socio-Economic Index for Areas
SES	Socio-Economic Status

EXECUTIVE SUMMARY

The SES score used by the Commonwealth Government in the calculation of funding for non-Government schools uses the calculated SES of the areas in which a student lives. SES scores are calculated at the SA1 level based on measures of family and household income, education and occupation. The values for each component are combined using a 1/3 weight for education; a 1/3 weight for occupation; and a 1/6 weight for household income and a 1/6 weight for family income. The values for each student are then used to calculate an average for the school.

The funding model is needs based and designed to take into account the parents' capacity to pay for their child's education. In the original specification of the SES index, occupation and education were included in the index to reduce the effect of variability in incomes for small areas.

Some of the critique of the SES score methodology has been that education and occupation are not relevant when assessing capacity to pay for a child's education – the main factor is income. Education and occupation should, therefore, not be included in the SES score. Education and occupation were originally included in the SES score calculation because it was theorised that income within small areas was variable over time, but there is limited evidence to support this.

This analysis attempts to answer the four research questions:

• How variable is personal income data at the small area level across years?

• Are income measures sensitive to economic changes and cycles?

• Are non-income variables still required to stabilise income dimensions of a capacity to contribute measure?

• Are there methods to use multiple years of income data to stabilise measures?

The first two questions are about the variability of income measures, while the second two questions are around ways to stabilise an income measure.

To answer the first two questions, we looked at the variability of incomes aggregated to the SA2 level using Personal Income Tax (PIT) data over 10 years; and to the SA1 level using Census data from 3 Census'. Variability in PIT data for each SA2 in each capital city/balance of State in Australia was identified by first rebasing incomes to the income for each capital city/balance of state in the base year (either 2005/06 or 2010/11) and calculating percentage changes in subsequent years compared to the base. This was then plotted for each capital city/balance of state as a time series to provide a graphical representation of the changes. Additionally, a basic regression with slope and intercept terms only was run for each SA2 and the root-mean-square errors (RMSE) were calculated, providing an easily interpretable measure of income variability relative to the overall trend.

The graphs also showed where 95% of the national values fell, for all years. Values falling outside these lines are outside the top and bottom 2.5% of values, so could be considered significant differences.[1]

The analysis of the Census data looked at quintile tables, showing changes in quintiles at the SA1 level in the SES score; and a score using household income only, both obtained from the Department. The Census data was also used to answer the third research question, by looking at the variability of the SES measure and an income only measure over three Census periods.

To answer the final research question, a 3 year moving average of the PIT data was used, and the same tests (graphs over time and RMSE) were used to summarise the variability.

In addition, for one period of the PIT data, median income was available rather than mean. Mean incomes are heavily influenced by extreme incomes, and are therefore known to be skewed. This means a mean, or average, is not a good summary measure as extreme values can influence it. Normally a median is used as a summary of incomes, and a median income is a lot more stable over time. Therefore, the summary figures for the PIT data (graphs and the RMSE) were also calculated for median incomes.

The PIT analysis was conducted at the SA2 level because this was the smallest area for which personal income tax data is publicly available. SA2s are also the smallest areas at which we could reasonably expect incomes to be stable from year to year. SA1 areas, the smallest areas for which most ABS data is publicly available, have an average population of approximately 400 individuals. Of these, only a fraction would be income earners; the remainder would be children, the elderly, unemployed or otherwise unable to work. With such low populations, small changes incomes as measured in dollars may represent large percentage changes.

The results showed that, for most areas, incomes remained relatively stable. The number of SA2's which had income changes within the 95% range of national income change was close to 90% in each time period looked at. In the first time period, 12% of SA2's had income change

[1] Note that this is not the same as assessing the difference as statistically significant – we haven't done tests of significance, we have merely looked at whether a value is inside or outside where 95% of the national values are.

outside the 95% national change; and in the second time period, 10% had changes outside the 95% national change. Areas with low populations of earners may experience large changes in average incomes over the period studied. Areas that have sufficiently large and diverse populations are unlikely to be affected by dramatic changes in income over short periods of time, however, some notable exceptions were observed.

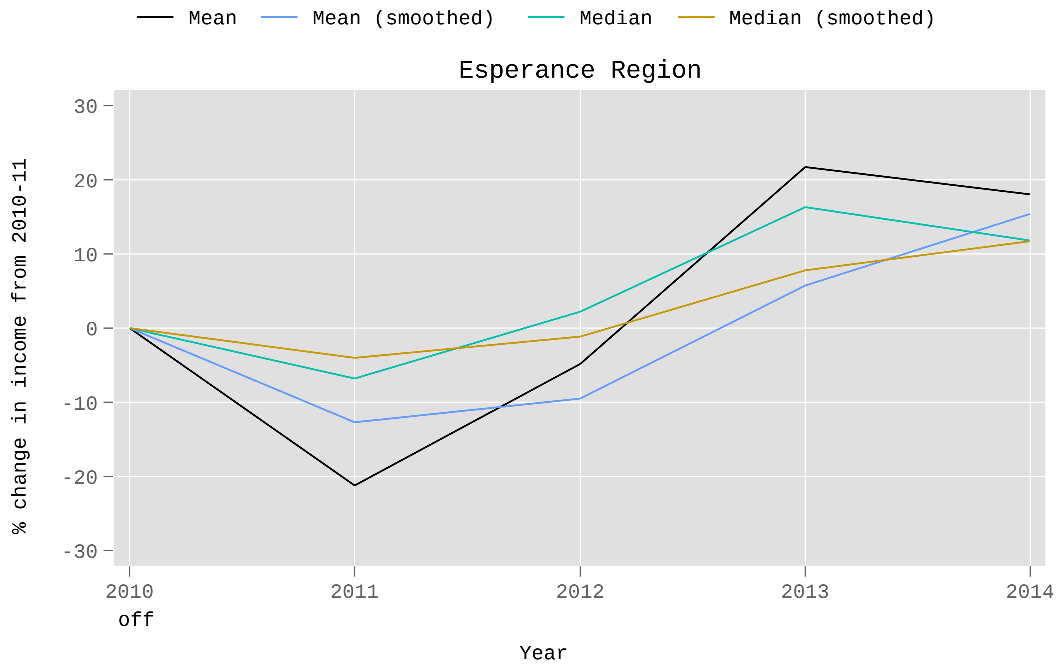
Some areas undergoing structural change, such as subdivisions and new building, gentrification and urban infill, or areas that are particularly dependent on specific industries, such as mining or agriculture, experienced relatively large changes in average incomes over relatively short periods of time.

The reasons for change will differ from place to place, and will likely be dependent on a combination of local planning policy and events affecting the broader economy. In some cases, incomes may rise and fall with the economy, in others they may rise or fall as the social makeup of the area changes. Examples of this might include a country town in decline where workers emigrate in search of better opportunities, or an inner suburb in which previously low-value housing is being demolished or renovated to accommodate wealthier households hoping to take advantage of the central location.

We found that median incomes were much more stable than mean incomes. This would suggest median income should be used in the SES measure. We also found that using a moving average reduced any variability further, but still reflected a change in incomes when this did occur in an area. The use of a moving average meant any income change was slowly brought into effect, allowing time for schools to respond to the changing situation.

An example of the relative stability of mean, median, and smoothed mean and median incomes from 2010/11 to 2014/15 is provided in Figure 1. The Esperance Region SA2 has a relatively variable income compared to most other SA2s, and this is especially visible when taking the mean income, which can vary by over 20%. If median incomes are taken instead, this variation is reduced considerably, but it is still able to capture the gradual increase in incomes observed over the period. A smoothed median income shows even less variability while still picking up the trend increase, as can be seen in Figure 1.

Figure 1 Change in income from 2010-11 to 2014-15 in Esperance Region SA2



Areas that do experience large changes in incomes that are not due to a low population base need to be assessed on a case by case basis making use of local knowledge about the area and what factors are likely to be driving changes in income. This is a fundamental limitation of any area based funding policy implemented where areas are not static. Most areas are relatively stable most of the time, but to ensure that those that do change, possibly dramatically and unexpectedly, are adequately funded, incomes must be monitored and any changes investigated. For one community in NSW which did have a large population, we identified a large increase in PIT income in one year only. Any anomalies like these can exist in administrative data, and they need to be identified, assessed and if need be, removed (or adjusted back to the trend) from any income dataset used to calculate funding as it would mean reduced funding in one year (the income increased more than 100% in one year, and returned to normal levels the next year).

In answer to the research questions asked:

- ***How variable is personal income data at the small area level across years?***

Income is generally very stable for most small areas across Australia. This is shown by the fact that approximately 88% of SA2s from 2005/06 to 2010/11 and 90% from 2010/11 to 2014/15 had changes in mean incomes within 95% of the national variation in mean incomes.

Instability occurred in areas of low population; and areas experiencing changes like drought, mining booms, urban renewal or new growth areas. Where mean income is variable, using a median income reduces this variability; and using a smoothed median income reduces it even further.

- ***Are income measures sensitive to economic changes and cycles?***

Yes – this research has found that outside factors will affect incomes, and can affect them dramatically. Drought, mining booms, new growth areas, business closures, etc, can all affect incomes. While we haven't been able to test in this analysis whether income in some areas is more affected by outside shocks (termed in the literature resilience), there is evidence in outside literature for higher capacity to adapt to external shocks in urban areas due to greater access to jobs (Vidyattama, Pearson, Tanton, & Mohanty, 2017), but this isn't just to do with incomes.

- ***Are non-income variables still required to stabilise income dimensions of a capacity to contribute measure?***

No. While income and education variables do provide some stability compared to an income only SES measure, there are other, more appropriate, ways of achieving the same result. In this paper, we have shown a number of ways of stabilising an income only measure without incorporating non-income variables.

- ***Are there methods to use multiple years of income data to stabilise measures?***

Yes. In this paper, we have shown that using median income; and using a three year moving average of median income stabilises the income measure. However, extreme changes in area level income can still occur, and need to be looked at each year to identify whether the extreme change is due to low population in the area, is a data quality issue, or is a real change caused by, for example, the closure of a large industry in the area.

1. INTRODUCTION

The SES score methodology used by the Commonwealth Government to calculate capacity to contribute for non-government schools uses the calculated SES of the areas in which a student lives. SES scores are calculated at the SA1 level and are based on measures of family and household income, education and occupation. The values for each component are combined using a 1/3 weight for education; a 1/3 weight for occupation; and a 1/6 weight for household income and a 1/6 weight for family income. The values for each student are then used to calculate an average for the school.

The school funding model is needs based and designed to take into account the parents' capacity to pay for their child's education by discounting the base funding amount. In the original specification of the SES index, occupation and education were included in the index to reduce the effect of variability in incomes for small areas (SA1).

Some of the critique of the SES score methodology has been that education and occupation are not relevant when assessing capacity to pay for a child's education – the main factor is income. Education and occupation should, therefore, not be included in the SES calculation. Education and occupation were originally included in the SES calculation because it was theorised that income within small areas was variable over time, but there is limited evidence to support this.

This report looks at the variability of incomes for areas over time, using 10 years of annual Personal Income Tax data at the SA2 level; as well as Census data at the SA1 level (the same data used to calculate the SA1 SES index). In particular, the datasets used are:

1) Two sets of Personal Income Tax (PIT) data at the SA2 level. These datasets cover the financial years 2005/06 to 2010/11, and 2010/11 to 2014/15. Due to changes in definitions[2], these datasets are not directly comparable and could not be merged. As such, each dataset was analysed separately.

---

[2] In 2005-10, Non-lodgers were included in the data; and the geographic boundaries were updated to ASGS 2016 - http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6524.0.55.002Explanatory%20Notes12011-2015?OpenDocument

2) Census family income data and the SES score at the SA1 level, the smallest area available for 2006, 2011 and 2016 Census data.

These two datasets used different indicators of income. The PIT analysis used a median or mean income for the area; whereas the SES score was an index calculated for an area from the proportion of people in the area earning above a certain household and family income cut-off; and below a certain household and family income cut-off. One of the limitations of Census data is that incomes are provided by the ABS in groups, so median and mean incomes are difficult to calculate without some imputation.

For the analysis of income stability in this paper, we mainly used family income, although some figures used household income (see, for example, Figure 75 here). This is because there was not much difference between the two in terms of income variability, so it was simpler to choose one for most of the analysis.

We need to be clear from the outset that all this analysis is area based, so for any particular area, averaging of incomes has already occurred across the area. This averaging can hide pockets of disadvantage in an area, and particularly affects areas like the ACT (Tanton, Vidyattama, & Mohanty, 2015). This averaging will be greater for larger areas, which is why we chose the SA1 area for the Census analysis, as this is the smallest area that income data for 2006, 2011 and 2016 is available.

Further, the incomes used in the analysis are incomes for all families and individuals in the area, not incomes of parents of non-Government school children. This is the same method used for the current SES score that the department calculates – an area based measure using all earners in the area.

2. RESEARCH QUESTIONS

This analysis looks at whether incomes for small areas in Australia vary over time, and what methods might reduce any variability. Specifically, the questions asked are:

- How variable is personal income data at the small area level across years?
- Are income measures sensitive to economic changes and cycles?

- Are non-income variables still required to stabilise income measures of capacity to contribute?
- Are there methods to use multiple years of income data to stabilise measures?

3. METHOD

The PIT data represent the average, annual, gross personal incomes for every SA2 in Australia. Data are collected by the Australian Tax Office (ATO) from tax records and are compiled and published by the ABS.

For the 2005/06 to 2010/11 data, only average individual income for each SA2 was available. It is known that incomes are heavily skewed, so an average is influenced by any extremely high incomes in an area. A better summary for incomes is median income, but this was not available for the 2005/06 to 2010/11 period. However, it was available for the 2010/11 to 2014/15 period, so for this time period, a mean and median income was used to test whether average incomes were more unstable than median incomes.

Incomes are expressed in current prices, and need to be adjusted to account for changes in the value of the dollar. This would normally be done using a deflator like average weekly earnings (AWE), but this was found to be unsuitable given the different sources of data. Incomes were adjusted to be in constant 2005/06 or 2010/11 dollars (depending on the series) using the average income in the capital city or balance of state. This meant that an SA2 with an income increasing at the same rate as incomes in their capital city/balance of state would show no variation in income.

Average Capital City/Balance of State incomes were used because the rate of income growth in capital cities like Melbourne or Sydney may differ from that outside capital cities. The incomes standardised to 2005/06 values were calculated as:

$$2005\_PIT_y^i = PIT_y^i * \frac{2005\_PIT}{PIT_y}$$

Where $2005\_PIT_y^i$ is the PIT for SA2 i in Year y in 2005/06 prices; PIT_y^i is the PIT for SA2 i in year y; 2005_PIT is the PIT in 2005/06 for the capital city/balance of state that SA2 i is in; and PIT_y is the PIT for year y for the capital city/balance of state that SA2 i is in. This calculation was done for the 15 capital city/balance of State areas in Australia (2 for each State and Territory except the ACT, which doesn't have a balance of state).

The percentage change from 2005/06 to each future year was then calculated and graphed for each SA2 in each Capital City/Balance of State. Assuming no real change in incomes (so if the income in the area grew at exactly the same rate as the average for that capital city/balance of state), all SA2s would show a 0 for every year. We expect some variation around this, as different areas experience different economic conditions and the deflator applied was based on the average for all SA2s in each capital city/balance of state. We do, however, expect the average for each SA2 to be close to 0.

Note that because the percentage change was from the base year rather than change from the previous year, the graphs are interpreted as total change over the period, i.e., if there is an increase in incomes in one year, and there is no change in the next year, the change line will show the next year as high also rather than returning to 0.

Plots were then made for each capital city/balance of state, and selected SA2s showing high variability were identified and labelled with their name, population range (i.e. the minimum and maximum populations over the period), and number of students enrolled in non-government schools. Populations were shown because low populations of tax payers in an SA2 are likely to contribute to greater instability of average incomes in that SA2. Population ranges (from the first year to the last), rather than mean populations, allowed high growth areas to be more easily identified. The number of students in the area was shown because if an area has high income variability, but a low student population, the variability may not be an issue – it will have a minimal impact on schools funding (unless all the students go to the one small school).

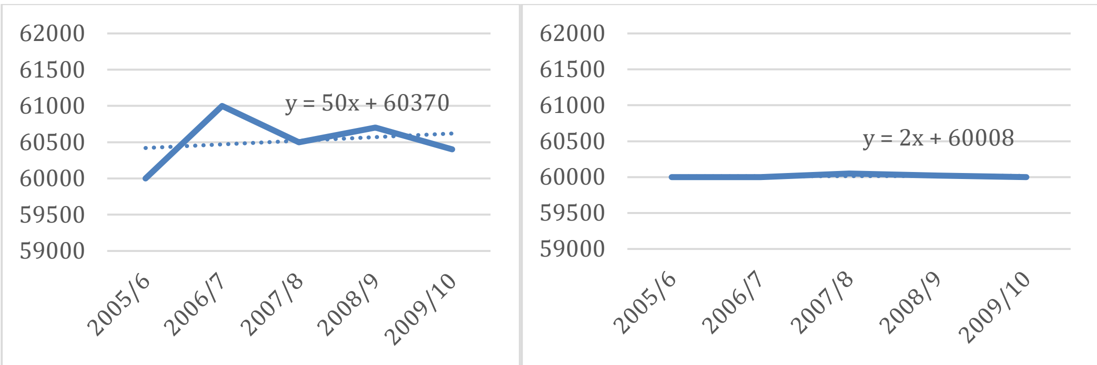
The Root Mean Squared Error (RMSE) is a standard statistical tool for measuring variability, and is normally used in regression models to estimate the variability of the estimated values around the true values of a dependent variable. In our case, we have used it to estimate the variability of actual income for each area around the trend income for the area over the time period of each PIT dataset. The RMSE represents the mean error around a regression line for incomes in each year.

$$RMSE_i = \sqrt{\frac{\sum_{y=1}^{n}(PIT_i^y - \widehat{PIT_i^y})^2}{n}}$$

Where RMSEi is the RMSE for SA2 i; PIT^y_i is the PIT income for SA2 i in year Y; $\widehat{PIT_i^y}$ is the estimate of PIT income from a simple regression model for SA2 i and year y with only an intercept and slope; i is the SA2; and n is the number of time periods (which is constant for all SA2's i). If the income in an SA2 doesn't change much or increases/decreases at a relatively consistent rate, then the RMSE will be low; whereas if incomes are inconsistent across years it will be high. An example for two fictional areas is shown in Figure 2.

In this figure, the RMSE for the first area is $144.64; and for the second area is $8.67. This is then better understood when expressed as a percentage of the average income, so in the first area, it is 0.24% of the average income; and in the second area it is 0.01%. The area with more variability in income shows a higher RMSE.

Figure 2: RMSE for 2 fictional areas



The RMSE was then plotted against the average population in the area using a scatter plot (see Figure 18 for an example here). This highlighted that areas with lower populations tended to have higher RMSE values.

We were also interested in the distribution of the RMSE's across Australia. For this, we calculated the mean, median and standard deviation of the RMSE values for all SA2's across Australia and show these in a table, as well as a frequency distribution (see Figure 19 for an example here).

Because the RMSE's can be difficult to interpret, we also calculated a frequency distribution of the total % change in PIT incomes from the first year to the final year for each time period (see Figure 20 for an example, here). Many areas show minimal change (-2 to 2%), but this graph does show that a very small percent are above or below 10% change.

Finally, we wanted to look at how variability in incomes changed over time, so a series of scatter plots was created which showed how the ranks changed each year. In these plots (Figure 21 is an example, here), the horizontal axis and vertical axis for each graph can be identified from the boxes with text in them – so, for example, the first graph in Figure 21 (so the second box reading from the top) is the graph with 2005 rank on the vertical axis; and 2006 rank on the horizontal axis. The first graph in the second row of Figure 21 shows the scatter plot with 2005 rank on the horizontal axis, and 2006 rank on the vertical axis. This is also summarised as a table of correlation coefficients (see Table 3 and 4 here).

Using all SA2's across Australia, we then calculated where 95% of values were; and then for each capital city/balance of state, we were able to plot these 95% lines on the graphs, and identify points that fell outside of these lines (so the top 2.5% and the bottom 2.5% of values in Australia). All the SA2's that were outside these lines are shown in Appendix 5, along with the minimum and maximum populations; and the change for each year, to show whether it is a consistent increase or decrease; or whether it just happened in one year.

All the graphs for mean incomes are shown in Appendix 1, (available here).

Finally, to answer the research question about whether there are methods to combine multiple years of income data to stabilise measures, we calculated a moving average of incomes using the PIT data for each period calculated using the formula:

$MA\_PIT^i = (1/6)*x^i_{t-2} + (1/3)*x^i_{t-1} + (1/2)*x^i_t$

Where MA_PIT^i is the 3 year moving average of PIT for SA2 i; and x^i_t is the PIT income for year t and SA2 i. This formula puts most weight on the year t (1/2), with lower weights for t-1 (1/3) and t-2 (1/6). The same graphs as used to analyse each PIT period were created for the smoothed income data, and the results are shown in Appendix 6 (see Appendix 6 here).

Results for all these same graphs using median incomes are shown in Appendix 7; and results for smoothed, median incomes are shown in Appendix 8. (All appendices are available here).

There are a number of issues with the PIT data that should be highlighted. The first is that it does not cover all taxpayers. Those receiving an income below the tax free threshold will be excluded; as will those people not paying tax. This will have a small impact on the mean and median used for our numbers.

For the Census data, only three data points for each region were available (compared to 10 using the PIT data), 2006, 2011 and 2016. The time between each observation was five years rather than one, however, the geographies available were much smaller (SA1 rather than SA2). This means the level of averaging will be less for these areas, so we would expect greater variability using these areas. However, the mean and median income, which was used for the PIT analysis, were not readily available from the Census data without using some imputation from a survey. For this analysis, we also wanted to use the data that the Department uses for the SES score, so the Department provided the SES score, and the underlying family income, household income, education and occupation indexes for each Census year. The family income data was then used to identify the variability in incomes at the SA1 level. The aim of this process was to identify whether an index using just income was more variable at the SA1 level than the SES index used in the school funding assessment.

Data from the 2006 census were for census collection districts (CD), rather than SA1, and were concorded to SA1 using a population-weighted concordance from the ABS to enable comparisons across years.

Note that the family and household income data were not means and medians, which the PIT data used. They were indexes based on the proportion of people in an area with incomes above and below certain cut-offs. The reason for this is that the Census has income data available in groups, so calculating a mean or median is difficult without some imputation using survey data (which can provide an income distribution within each of the Census income groups). The SES score therefore uses a proportion of households or families with incomes above; and below certain cutoffs. This is similar to the way SEIFA incorporates income data into an index of disadvantage.

Because the index were ordinal values, we used quintile tables, which show the quintile that the area was in for one index, compared to the quintile it was in for a second index. For this analysis, we used non-government school children weighted quintiles, so each quintile contained an equal number of non-government school children. These number of non-government school children in each area was provided by the Department. These quintile tables are used by NATSEM regularly to compare ordinal indexes, for example, the child social exclusion and SEIFA index.

To compare indexes, we looked at quintile tables comparing SES to a household income only index in 2006, 2011 and 2016; and then looked at the variability over time for the SES index from 2006 to 2011, and 2011 to 2016; and the Family Income index from 2006 to 2011, and 2011 to 2016. This gives an idea of the difference if an income only index had been used rather than the SES index for each Census year; and also, the variability of the family income and SES indexes over time. The quintile tables are shown in Appendix 2.

We also looked at a time series of change in family income at the SA1 level, similar to the analysis with PIT data, but for 3 time periods rather than 5. This means that the variability will be greater, as each data point shows 5 years of change, not one year. These graphs are shown in Appendix 3.

In addition to the quintile analyses, we also explored changes in each domain index at the SA1 level: SES, family income, household income, occupation, and education. To do this we calculated the change in rank from 2006 to 2011 and from 2011 to 2016 in absolute values (e.g. -1 is equivalent to 1). We then summed these to show the total absolute change in rank over the three periods.

Rank values were plotted for each SA1 in each year compared to each other year as a scatterplot matrix, and these are shown in Appendix 4 here. These graphs are interpreted in the same way that the series of scatterplots for the PIT data are interpreted. Distributions of rank changes were also plotted in Appendix 4 in Figure 78 and Figure 79 (available here).

Finally, we chose some case study areas, and looked at the change in PIT data for these areas, while also looking at the change in the population. For one case study area, we have looked at the mean, median and smoothed median to show the effect of using a smoothed median income for an area where mean income is unstable. These case study areas are Emerald as a mining area in WA; Esperance as a mining area in WA with unstable income; Nhulunbuy as a mining area in another State (the Northern Territory); Crace as a high growth area in the ACT; and Redfern-Chippendale as an area of gentrification in Sydney.

4. RESULTS

PIT ANALYSIS

The results from the PIT analysis are shown in full in Appendix 1. This section discusses the overall results.

Overall, it can be seen that the vast majority of areas had stable incomes that did not change by more than a few percent, with 88% (2005/06 to 2010/11) or 90% (2010/11 to 2014/15) of SA2s having all changes in income falling within the 95% range of observed national change. There were, however, some areas that had highly unstable average incomes. These tended to be those areas with low populations of earners. This can be seen most clearly in Figure 18 and Figure 37 here, where the RMSE is plotted against the mean population for each area and shows that as population increases the mean error decreases sharply and generally remains low. The plot does, however, show that there are some areas with larger populations and highly unstable incomes. These areas are likely to be experiencing some sort of substantive change in their structure. This could include gentrification and new building or the beginning or end of a boom or bust cycle (e.g. in mining or agriculture). Areas with higher variability and higher populations also tended to be outside the capital cities, where economies are typically less diversified and more vulnerable to external shocks.

From 2005/06 to 2010/11, there were no areas with high variability (ie, outside the 95% national lines) in Hobart. There were more areas with high variability in the 2010/11 to 2014/15 period, although Hobart and Adelaide had no areas outside the 95% national lines.

Looking at the areas that were highly variable in the 2005/06 to 2010-11 period, many were affected by the 2001 – 2009 drought. Narrabri and Grenfell in particular saw consistently lower incomes over this period. Condobolin and Narrabri saw a sharp increase in incomes in 2010/11, which was the year the drought broke.

Mildura in Victoria saw a sharp increase in mean income in 2007/08, but this settled back to normal levels in 2008/09, suggesting it may be a one-off event or an artefact of the data.

In Brisbane, Eagle Farm and Brisbane Airport both had low populations, which would contribute to their instability. Kilroy had one year where incomes were higher than normal, but this returned in the next year, suggesting an anomaly in this year.

For areas outside Brisbane, a number experienced consistent change in incomes – for example, Glass House Mountains saw a reduction in 2006/07, and then incomes stayed at this lower point, suggesting 2005/06 may have been an anomaly. Incomes in Northern Peninsula are consistently growing above other areas in rural Queensland, with a reasonable population, suggesting some structural change in this area.

All the instability in the Adelaide SA2's can be attributed to low population areas. Outside Adelaide, Franklin Harbour and Le Hunte both experience consistently increasing incomes, suggesting some structural change in these areas, possibly driven by mining and ports in these two areas.

In Perth, Malaga is an industrial precinct, with low population and Dawsville is a relatively new area South of Perth, and experienced very high population growth over this period, contributing to the high levels of income growth. Outside Perth, the average incomes in many areas increased rapidly in 2007/08 and 2008/09, possibly driven by the mining boom over this period; but many of the areas were also agricultural (eg, Kulin is in the wheatbelt), so subject to income variation due to changing weather patterns.

SA2's in Hobart and the balance of Tasmania experienced very low variability in incomes over this period.

Darwin had only one SA2 that experienced a notable decrease in income, Darwin City, but this seemed to happen in 2006/07 and 2007/08, and it has been consistent since then. Areas in the NT outside Darwin experienced some instability in incomes, mainly due to low population counts.

Areas in the ACT with unstable incomes were Crace, a new high growth area in Gungahlin; and Acton, which experienced a drop in incomes in 2006/07 and 2007/08, but was fairly consistent after this. There were many new student residences for the ANU being built in Acton over this period, which may have contributed to the lower average income.

Looking at the graph of RMSE and population (Figure 18, available here), it is clear that low population areas have higher RMSE. However, there are some standout areas like Mosman and Double Bay that are high population and high variability in incomes (increases in Mosman). This could be driven by demographic change in these areas, as older and lower income populations leave, and younger families with higher incomes buy into these areas.

Looking at the figures from 2010/11 to 2014/15, in Sydney it can be seen that the areas with high levels of variability were generally those with low populations. Chullora was a high growth area, and Redfern-Chippendale showed an increase in incomes in 2014/15, but this is also an area of urban renewal.

Outside Sydney, Scone showed increasing incomes, although the value in 2013/14 seems to be an anomaly requiring further investigation. Walgett-Lightning ridge showed decrease in income, possibly due to the mining boom ending around this period.

In Melbourne, two of the areas with high instability were low population areas (Melbourne Airport and Flemington Racecourse); and one (Wallert) seems to be a high growth area, suggesting a new development and an influx of people, contributing to the high variability in incomes.

Outside Melbourne, Alps West had a low population, and Yarram in South Gippsland experienced high income growth in 2013/14 which stayed high in 2014/15. Yarram is a dairy farming area which may affect incomes in the area.

In Brisbane, unstable areas were all low population, while in the rest of Queensland, Gladstone was the only area with a larger population and increasing incomes, possibly due to the tail end of the mining boom, which was becoming a bust in many other areas around this time.

In Adelaide, all SA2's had very small changes in income, except Lonsdale, which was a low population area. In the rest of SA, Le Hunte-Elliston experienced lower incomes. This SA2 experienced much higher incomes in the 2005/06 to 2010/11 period, so this might signify the end of the mining boom for this port area, and a return to more normal incomes.

In Perth, income variability was likely due to low populations, and one area of high growth (North Coogee), which looks like a new development in Perth. There was some instability in SA2's in WA outside Perth, as there was in the 2005/06 to 2010/11 period, and the areas were the same (Kulin, Esperance), suggesting that this variability in incomes is due to either mining or farming. Kulin is interesting, as it is in the wheat-belt, but incomes have been increasing consistently over the second period, so it isn't up and down, as we might expect to see in an agricultural area; it has been up over the 5 years. This was different to the 2005/06 to 2010/11 period, where it was up and down.

Hobart and other Tasmanian SA2s showed very little variability in incomes.

Darwin had two areas, but both were showing a changing trend, rather than variability. Brinkin was trending down, and Berrimah up (although it then dropped off from 2013/14). The rest of the NT did show variability in incomes, mainly due to low population.

In the ACT, the two SA2's with varying incomes were both low population areas (Scrivener and Majura).

Figure 20 (available here) summarises the distribution of % change in incomes from 2005/06 to 2010/11 for all SA2's across Australia. It can be seen that about 37% of SA2's experience a change in incomes between -2 to +2% change (the first bar), while most SA2's are between the -6% and +6% bars.

The scatterplot of RMSE's to population for 2010/11 to 2014/15 again showed higher RMSE's for lower population, as expected. Again, Scone came up as an odd SA2, with a reasonable population and very high RMSE, but we suspect this is to do with the one odd income in 2013/14, as discussed above.

Figure 39 (available here) summarises the distribution of % change in incomes from 2010/11 to 2014/15 for all SA2's across Australia. It can be seen that similar to the previous time period, about 36% of SA2's experience a change in incomes between -2 to +2% change (the first bar), while most SA2's are between the -6% and +6% bars.

Overall, the average RMSE calculated for the years 2005/06 to 2010/11 for the PIT incomes for all SA2's across Australia was about $694, or about 1.7% of the average income of $39,700 (in 2005/06 dollars). For the years 2010/11 to 2014/15, the average RMSE was $926 with an average income of $51,300 in 2010/11 dollars (so 1.8% of income) (Table 2). The distribution of the RMSEs for 2005/06 to 2010/11 is shown in Figure 19 and the results for 2010/11 to 2014/15 are shown in Figure 38. It can be seen that the distribution is highly skewed, with the majority of SA2's having RMSE's of less than $1000 per year (2.5% of average income in 205/06 and 1.9% in 2010/11). In both years, nearly 85% of areas had an RMSE of less than $1000 per year.

Using the median value for the second time period, it can be seen in Table 2 that the RMSE using a median income is much lower than the RMSE using a mean income. Using a mean income, the average RMSE for all SA2's was $926.21, whereas using a median it was $616.43. The variability around this mean was also much lower for median income rather than mean, with the standard deviation of the mean RMSE being $1,663.34 compared to $794.73 for median income. The full results (line graphs for the 2010/11 to 2014/15 period for all SA2's in each capital city/balance of state; distribution of RMSE's for all SA2's; and scatterplot of population and RMSE) for median income is shown in Appendix 7, and the results for a smoothed median income are shown in Appendix 8 (all appendices are available here).

The plots in Appendix 8 show that the SA2 income variability is much lower using a smoothed median income compared to the original unsmoothed mean income. There are still some areas experiencing income variability, but they are not as extreme; and are generally the low population areas (see Figure 149 here). Comparing the RMSE's, the distribution using a smoothed median income (see Figure 150 here) is much tighter than using mean income (see Figure 38 here).

Table 1 Comparison of RMSEs using median and mean incomes for all SA2's, 2005/06 to 2010/11

	Mean	
	RMSE	**RMSE (smoothed)**
Obs.	2127	2127
Mean	693.84	367.42
Std. Dev.	834.03	466.2
Median	450.35	239.69

Table 2 Comparison of RMSEs using median and mean incomes for all SA2's, 2010/11 to 2014/15

	Median		**Mean**	
	RMSE	**RMSE (smoothed)**	**RMSE**	**RMSE (smoothed)**
Obs.	2193	2193	2193	2193
Mean	616.43	310.63	926.21	464.86
Std. Dev.	794.73	384.48	1663.34	797.81
Median	425.65	212.19	525.40	263.91

CENSUS DATA ANALYSIS

The results from the analysis of the Census 2006, 2011 and 2016 are shown in Appendix 2 (see the following link for all appendices). The results comparing family income and SES for each Census year (Tables 3 – 6) show a strong diagonal, with some students in the quintile one away from the main diagonal; but not many beyond that. As an example, looking at the first cell in Table 7 (reproduced below), if family income only had been used rather than SES, then about 75% of SA1's would be in the same quintile for SES and family income; 21% are currently in quintile 2 using SES but would be in quintile 1 using family income only, meaning more funding; and 22% are currently in quintile 1 using SES, and would move to quintile 2 if family income only was used.

Table 5 (reproduced from appendix): SES vs Income quintiles, 2016 showing percent and count of students

	2016 Family Income Quintiles					
2016 SES Quintiles	**1**	**2**	**3**	**4**	**5**	**Total**
1	74.58 (157962)	22.17 (46961)	2.98 (6320)	0.26 (543)	0.01 (23)	100 (211809)
2	21.48 (45475)	49.28 (104346)	25.14 (53241)	3.89 (8230)	0.21 (453)	100 (211745)
3	4.01 (8504)	23.6 (49995)	47.45 (100531)	23.54 (49872)	1.4 (2959)	100 (211861)
4	0.46 (970)	4.3 (9110)	22.76 (48193)	55.71 (117961)	16.77 (35504)	100 (211738)
5	0.01 (24)	0.17 (358)	1.58 (3336)	16.78 (35529)	81.46 (172476)	100 (211723)

Note: Totals are not the same in each quintile because an area based analysis means inexact cut-offs for the quintiles.

Generally, there is more movement outside the extreme quintiles (1 and 5), so in Table 7, the 1-1 figure is 74.58% while the 2-2 figure is 49.28%, with another (22.17 + 23.6) = 45.77% moving one quintile away, and the other (4.3 + 0.17) = 4.47% moving more than 1 quintile. This is because there is only one direction to move for quintiles 1 and 5, whereas quintiles 2 – 4 have two directions to move.

In all years, the number of areas moving by more than 1 quintile is less than 10%; the proportion moving more than 2 quintiles is 1 – 2%; and the proportion moving more than 3 quintiles (which can only be done from quintiles 1 and 5) is rounded to 0%.

We also looked at the change over time (2006 – 2011 and 2011 – 2016) of family income and SES indexes. The results are shown in Tables 3 – 9 of Appendix 2 (available here). Looking at Table 8, which shows family income in 2006 and 2011, it can be seen that 69% of SA1's that were in quintile 1 in 2006 stayed in quintile 1 in 2011; 24% of SA1's moved from quintile 2 in 2006 to quintile 1 in 2011; and 24% moved from quintile 1 in 2006 to quintile 2 in 2011. In all cases, less than 10% moved more than 1 quintile of income.

Looking at SES, this was much more stable over time, with less than 1% of SA1's moving more than 1 quintile over the 5 years from 2011 to 2016; and about 1% from 2006 – 2011.

What this analysis has shown is that at the SA1 level, there is more variability in income at the SA1 level than using SES. This is not surprising, as the education and occupation in the SES index is fairly constant over time, so would average out any variability in income. Using family income only, from 2006 to 2011, 69% of SA1's in the first quintile of family income in 2006 would be in the same quintile in 2011; whereas using SES, 79% would be. However, most of the movement is one quintile away.

Looking at the comparison between family income and SES within each year, there would be changes, and for about 5% of areas, the changes are beyond 1 quintile.

Looking at the maps of changes, it can be seen that areas with high levels of change in family income (see Figure 41 here) are in low population areas of regional and remote Queensland. Variability in SES (see Figure 42 here) is not as great, but again, remote areas experience grater change, due to low populations.

The scatterplots of changes in Appendix 4 confirmed the results from the quintile tables, which was that SES was more stable over time than income only (family or household). While the scatter plots are difficult to interpret, the graphs of the distribution of change in index scores Figure 78 and Figure 79 (available here) show that SES has a lower change in values than the income, education or occupation indexes. The second lowest was the Education index, then Household Income, Occupation and finally Family Income.

CASE STUDY ANALYSIS

This analysis chose a number of areas, and looked at the variability for these areas. The specific graphs for these areas are shown in Appendix 9 (Appendix 9, is available here). It can be seen that the income for Emerald was stable for 2005/05 to 2010/11; but there was a lot more instability in the second period of PIT data. The number of earners in Emerald was increasing from 2005/06 to 2010/11, but then decreasing from 2010/11 to 2014/15 (remembering that the definition of income earners changed between these two periods, so the second period has a much higher number).

In Nhulunbuy, the population of earners was unstable in the 2005/06 to 2010/11 period; but on a decrease in the 2010/11 to 2014/15 period (again, the changed definition affected this population). The incomes over this period also showed instability.

In Redfern-Chippendale, an inner Sydney suburb subject to gentrification, the number of earners was growing but stable (except 2007/08), however income was very unstable for the first period; and then seemed to be stagnant in the first four years of the second period, before increasing more than 30% in one year, from $56,000 to $74,000, suggesting a significant change in this year. Crace (a new area in the ACT) showed a similar pattern, with a very constant income growth in the first 5 year period while it was still being built, but then instability in the second 5 year period. The final income point in 2014/15 is lower than other years, but the scale on the Y axis means this is only a $1,000 difference from 2013/14.

The final case study was looking at Esperance in WA, and rather than looking at income instability for this, we looked at mean; smoothed mean; median; and smoothed median incomes. The results for this case study have already been shown in the executive summary.

5. CONCLUSIONS

This analysis has found that the main reason for instability in incomes is low population, with secondary external influences possibly including drought in regional areas; changes in incomes due to mining and farming; urban renewal in a few areas in cities; and new suburbs in cities. One area in the ACT (Acton) saw rapid income change due to demographic change as student accommodation for the ANU was built. However, it can be seen from the graphs that most SA2's across Australia showed very stable incomes. Using a more objective analysis, the number of SA2's which had income changes within the 95% range of national income change was close to 90% in each time period looked at. In the first time period, 12% of SA2's had income change outside the 95% national change; and in the second time period, 10% had changes outside the 95% national change. The areas that did fall outside the 95% range tended to have lower populations, with a mean of 2,510 compared to 4,792 for areas that did not in 2005/06 – 2010/11, and 2,820 compared to 5,936 in 2010/11 to 2014/15.

Using a Root Mean Squared Error provides a summary statistic on the variability of each area. In both periods looked at, the RMSE for nearly 85% of areas was less than $1,000 per year, or around 2 – 2.5% of the average income of the area. So most areas in Australia have relatively stable incomes. Of the 15% of SA2s that had an RMSE of more than $1,000 (2% of income), 18% of these had a population of less than 1,000 people in 2005/06 to 2010/11, and 15% in 2010/11 to 2014/15.

Using Census data at the SA1 level, and using the SES and family income data from the Department, we found that there was more variability in family income over time compared to SES. The proportion of SA1's moving more than 1 quintile in income between Census's was about 5 - 8% each year, whereas for SES it was 1 – 3%. This is not surprising given the SES measure has education and occupation which averages out any variability in incomes.

Looking at if an income measure had been used for each Census year rather than SES, the amount of variability showed that 5 – 6% of SA1's would change more than 1 quintile each Census year if family income were used instead of SES.

In conclusion, we have found that generally income is stable across time, particularly if median incomes are used. There is some instability in incomes across small areas, but most of it is due to low population areas; or economic and demographic change in an area, so it can be explained. If there is any instability, the impact of it can be reduced using a median, and using a moving average. An example was an odd mean income for Scone in 2013/14. Using a median income removed this odd value, suggesting it was due to an extreme income which does not have as much effect on a median as it does on a mean.

Using a smoothed median income, any variability in income that is left happens slowly over time, and can usually be attributed to external events, like localised economic booms or busts, or local demographic change. This means there is no need to use non-income variables to stabilise income dimensions of a measure of capacity to contribute.

REFERENCES

Tanton, R., Vidyattama, Y., & Mohanty, I. (2015). Disadvantage in the australian capital territory. *Policy Studies*, *36*(1). https://doi.org/10.1080/01442872.2014.981062

Vidyattama, Y., Pearson, L. J., Tanton, R., & Mohanty, I. (2017). Assessing adaptive capacity during the drought period in the Murray–Darling Basin. *Asia-Pacific Journal of Regional Science*. https://doi.org/10.1007/s41685-017-0039-z