View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

ISSN 1440-771X

MONASH UNIVERSITY
AUSTRALIA

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Statistical Inference on Changes in Income Inequality in Australia

George Athanasopoulos and Farshid Vahid

Working Paper 9/2002

Statistical Inference on Changes in Income Inequality in Australia

George Athanasopoulos*
Department of Econometrics and Business Statistics
Monash University Clayton, Victoria 3800.
AUSTRALIA

Farshid Vahid
Department of Econometrics and Business Statistics
Monash University, Clayton, Victoria 3800.
AUSTRALIA

August 20, 2002

Abstract

This paper studies the changes in income inequality of individuals in Australia between 1986 and 1999. Individuals are divided into various subgroups along several dimensions, such as region of residence, age, employment status etc. The changes in inequality over time, between and within the various subgroups is studied, and the bootstrap method is used to establish whether these changes are *statistically significant*.

Keywords: Income Inequality, Gini Coefficient, Theil Inequality Measure, Bootstrap.

JEL classification: D31, J10, C19.

*Corresponding author. E-mail: George.Athanasopoulos@BusEco.monash.edu.au

I Introduction

It has been documented that income inequality in most developed countries has increased during the past two decades[1]. Australia is not an exception[2]. In this paper we study the changes in the inequality of income of individuals in Australia between 1986 to 1999. We divide the data into subgroups along several dimensions and determine if inequality has increased *significantly* over time, and if so, has the increase been due to an increase in inequality between these subgroups or within each subgroup. We use the bootstrap method to establish the statistical significance of these changes.

Most of the previous research in Australia has concentrated on *earnings* inequality (see Borland 1999 for a survey). These studies fit in the literature on labour market dynamics, in which the main purpose is to study the distribution of returns to education, tenure and skills. Their results are only indirectly indicative of the evolution of the distribution of *income* in the society. However, for social welfare considerations, it seems inappropriate to focus on wage earners only. In particular, with the ever increasing proportion of aged population, understanding the changes in inequality in this subgroup is quite an important issue for policy makers.

In the first part of this study, we use the 1986, 1991 and 1996 one percent census data sets. These data sets, in particular the 1991 and 1996 ones, have more geographical details than any other available Australian data set. This allows us to study the changes in inequality in different geographical areas. The census data sets are also more "representative", in the sense that they are independent draws from the Australian population, and this makes the estimation and inference of inequality parameters based on these data sets relatively straightforward.

In the second part of the study, we use the 1993/94 and 1998/99 Household Expenditure Survey data sets. These data sets report *actual weekly* income[3] from all sources for all

[1] See, for example, Gottschalk and Smeeding (1997) and Bernstein et al (2001).

[2] See "Income Distribution 1999-2000" published by the Australian Bureau of Statistics.

[3] Weekly income can be quite a noisy indicator of income for some individuals. Unfortunately, annual income is only reported in the 98/99 survey.

individuals. However, the sample is chosen through stratified sampling procedures, in which different sample points may not have equal probability of being selected. Fortunately, these sampling "weights" are reported for each sample point, and we exploit these weights in the estimation of inequality measures. We also account for these weights in our bootstrap procedure as suggested by Biewen (2002). Using weekly incomes, we investigate how income inequality has changed for different age and occupation groups in Australia.

We use "individual" rather than "household" or "family" as our income earning unit throughout our analysis. Although using individuals brings in higher variability in incomes than using household or family income, it avoids the necessity of adjusting for the size and the age composition of the households. Although everyone agrees that family income must be adjusted to reflect the size and age composition of the family, there is no universally accepted method for doing this. Several methods – known as "equivalence scales" – for adjusting family income for scale economies have been suggested[4], and the evidence on the effect of the choice of alternative equivalence scales on the measurement of inequality has been mixed[5]. In this paper we look at individuals of at least 15 years of age, and we ignore income sharing in the family unit. Although this may give an exaggerated impression of inequality in the society, we believe that it will not affect our objective of studying the changes in inequality over time[6].

The outline of the paper is as follows. In section *II* we present the measures of inequality that we use in this paper, namely the Gini coefficient and Theil's inequality measure. A brief explanation of the bootstrap procedure from independent and stratified samples is provided in section *III*. We discuss the particulars of our data sets and present our results in section *IV*. Section *V* concludes.

[4] See Atkinson (1983, pp.48-53), Cowell (1995 p.99) and Coutler et. al (1992). Harding (1994) uses equivalence scales to study inequality in Australia.

[5] See Gottschalk and Smeeding (1997).

[6] Using the US data, Karoly and Burtless (1995) present evidence that indicates that equivalence scales may affect the level of measured inequality, but not its trend.

II Measures of Income Inequality

We consider two measures of inequality in this paper, the Gini coefficient (G) and the Theil measure (T). Both of these satisfy the three basic criteria of acceptable inequality measures (see Sen 1991), which are:

1. Invariance to unit of measurement of income;
2. Invariance to replication of population, i.e. if population size is doubled by adding an exact replica of every individual to the population, the inequality measure does not change;
3. Compliance with the Pigou-Dalton principle of transfers, which requires the inequality measure to decrease (or at least not to increase) any time that income is redistributed from a richer person to a poorer person, and vice versa.

We consider Gini because it is well known and we consider Theil because it is additively decomposable and satisfies a stronger version of the principle of transfers. We discuss these briefly below. Readers who are familiar with these measures may proceed to Section *III*.

(i) Gini Coefficient - (G)

The Gini coefficient attributed to Gini(1912), is probably the most widely used inequality measure. It is best understood as a measure of the area between the income Lorenz curve and the 45 degree line (the line of absolute equality). More precisely, the Gini coefficient is the ratio $\alpha/(\alpha+\beta)$ where α and β are the areas of the regions marked by these letters in Figure 1.

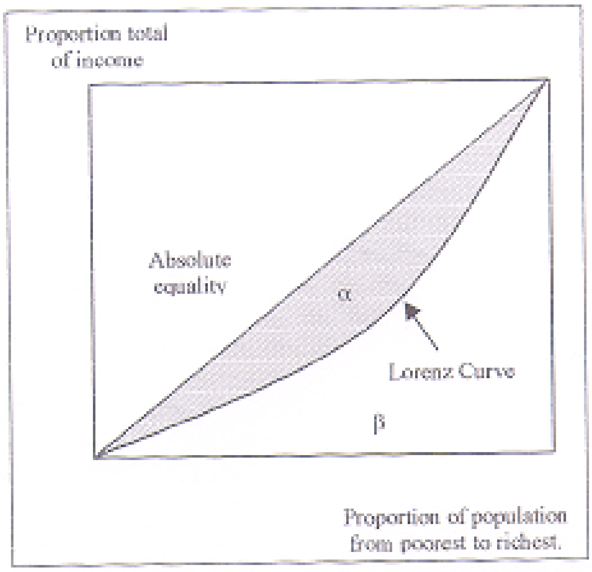

Figure 1: Lorenz Curve

Algebraically, the Gini coefficient is one half of the mean of the absolute values of differences between all pairs of incomes relative to the mean income, i.e.,

$$G = \frac{1}{2n^2\overline{y}} \sum_{i=1}^{n}\sum_{j=1}^{n} \mid y_i - y_j \mid, \tag{1}$$

where y_i is the income of individual i. The Gini coefficient can alternatively be calculated as

$$G = 1 + \left(\frac{1}{n}\right) - \left(\frac{2}{n^2\overline{y}}\right) [y_1 + 2y_2 + ... + ny_n] \quad \text{where} \quad y_1 \geq y_2 \geq .... \geq y_n. \tag{2}$$

The second formulation requires sorting the data first, but involves a single summation only. Since the Gini coefficient is typically calculated using samples with thousands of observations, it is worth noting that the second formulation is computationally much faster[7] than (1).

(ii) Theil's measure of inequality - (T)

The alternative measure of income inequality employed in this paper is Theil's (1967) measure of inequality

$$T = \sum_{i=1}^{n} s_i \ln(ns_i), \tag{3}$$

[7] This assessment is based on using the function sortc in GAUSS to sort the data.

where $s_i = y_i/(n\overline{y})$ (i.e. the share of individual i of the total income). An alternative formulation of the Theil measure that makes its relation to the entropy of the income distribution transparent is

$$T = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\overline{y}}\ln\left(\frac{y_i}{\overline{y}}\right). \tag{4}$$

One advantage of Theil's measure over the Gini coefficient is that it not only satisfies the Pigou-Dalton principle of transfers but it also complies with what Cowell (1995) refers to as the "Strong Principle of Transfers". This principle requires that the measure of inequality be more sensitive to transfers of income in the lower tail of the income distribution than it is to the transfers in the upper tail. For instance, a \$1 transfer from the poorest person to the second poorest person and a \$1 transfer from the second richest to the richest person both increase the Gini coefficient by the exact same amount, whereas the first transfer increases the Theil measure more than the latter does.

Another great advantage of Theil's inequality measure is that it is additively decomposable. This makes the Theil measure particularly attractive for econometric analysis of income inequality conditional on population characteristics. For example, for any significant grouping of the population (such as grouping by age, region of residence, occupation, etc.), the Theil measure of inequality in the entire population can be additively decomposed to inequality between subgroups (BT) and an appropriately weighted average of inequality within each group (WT).

$$T = BT + WT = \sum_{j=1}^{J} s_j^* \ln\left(\frac{n}{n_j}s_j^*\right) + \sum_{j=1}^{J} s_j^* \sum_{i=1}^{n_j} s_i^j \ln(n_j s_i^j), \tag{5}$$

where s_j^* is the share of the total income enjoyed by subgroup j, s_i^j is the share of the total income in group j enjoyed by individual i, and n_j is the number of people in group j.

(iii) Estimators for Gini and Theil coefficients

The Gini and Theil coefficients are often estimated on the basis of a sample drawn from the population. When the sample is collected via independent draws from the parent population, then the sample Gini and Theil coefficients will be method of moments estimators for the population Gini and Theil.

In many surveys, however, the subjects are not chosen completely randomly from the parent population, but they are chosen through stratified sampling. That is, the population is stratified according to some characteristics, and the sample is put together from random draws from each of these strata. Hence, two observations chosen from two different strata containing different numbers of people, would not have had equal chance of being selected in a random sample. In these cases, a "weight" is reported for each observation in the data set. These weights are proportional to the population of the strata which the sample observations are drawn from. In such cases, the estimates of the population Gini and Theil coefficients are calculated using the weighted Gini and Theil formulae:

$$G_w = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} \mid y_i - y_j \mid p_i p_j}{2\overline{y}_w}, \tag{6}$$

$$T_w = \sum_{i=1}^{n} p_i \left(\frac{y_i}{\overline{y}_w}\right) \ln\left(\frac{y_i}{\overline{y}_w}\right), \tag{7}$$

where $\overline{y}_w$ is the weighted average of incomes in the sample, and p_i and p_j are the normalized weights of observations i and j, where normalization is done to ensure that the weights sum to one over the entire sample. A computationally more efficient formulation of the weighted Gini is

$$G_w = 1 + \frac{\sum_{j=1}^{n} y_j p_j^2}{\sum_{j=1}^{n} y_j p_j} - \frac{2}{\sum_{j=1}^{n} y_j p_j} \sum_{j=1}^{n} \left(y_j p_j \sum_{i=1}^{j} p_i \right), \tag{8}$$

and the decomposition of weighted estimator of Theil coefficient to between and within components is given by

$$T_w = BT_w + WT_w = \sum_{j=1}^{J} P_j \left(\frac{\overline{y}_w^j}{\overline{y}_w}\right) \ln\left(\frac{\overline{y}_w^j}{\overline{y}_w}\right) + \sum_{j=1}^{J} P_j \left(\frac{\overline{y}_w^j}{\overline{y}_w}\right) T_w^j, \tag{9}$$

where P_j is the sum of weights of all observations in group j, $\overline{y}_w^j$ is the weighted average income of group j, and T_w^j is the weighted estimate of the Theil measure of inequality within group j.

III Bootstrap Methodology

There is an evident lack of statistical inference in the literature on measurement of income inequality. These studies typically report the point estimates of different measures of inequality and how these measures have been changing over time, but they rarely ask the question of whether these changes are statistically significant or not. Asymptotic inference seems particularly apt given that one prefers not to impose any functional form for the distribution of income, and because inequality measures are usually estimated using samples of thousands of observations. However, as Mills and Zandvakili (1997) point out, the rate of convergence of these complicated nonlinear functions may be slow and the confidence intervals based on the asymptotic distribution may extend outside the bounds of these measures. Moreover, when studying the components of a decomposable measure of inequality, the derivation of asymptotic covariance matrix of the estimators of various components can be challenging. For these reasons, bootstrap methodology (Efron 1979) is becoming popular for statistical inference on trend in inequality. At the least, bootstrap method can be viewed as substituting computer time for human time in deriving the asymptotic standard errors and confidence intervals for measures of inequality.

Mills and Zandvakili (1997) use bootstrap to study the changes in inequality in the United States. However, they ignore the fact that the survey data that they base their analysis on are not independent draws from the population, i.e., they do not take account of the "observation weight" reported for each observation. In a recent paper, Biewen (2002) has studied the bootstrap method for inference on inequality measures based on non-random samples. Since we use both the 1% census data and the household expenditure survey data[8], we briefly explain the bootstrap methodology for data sets created by both random and stratified sampling procedures.

(i) Bootstrapping from samples of independent observations

Suppose we have a random sample of size n, which is drawn from a completely unspecified probability distribution F. Let $\widehat{I}$ denote the point estimate of the inequality measure of interest based on this sample. Denote the empirical distribution of the sample by $\widehat{F}$, which is formed by attaching probability $1/n$ to each observation y_i for $i = 1$ to n. From $\widehat{F}$, B

[8] More details on the data are provided in Sections *IV(i)* and *IV(ii)*.

bootstrap samples of size n are drawn with replacement

$$\widehat{F} \longrightarrow \{y_1^{*b}, y_2^{*b}, ..., y_n^{*b}\} \qquad \text{for} \quad b = 1, 2, ..., B, \tag{10}$$

and the inequality measure of interest I^{*b} is calculated for each sample. The estimated standard error of I^{*b} calculated from the sample of B observations is a consistent estimator of the standard error of $\widehat{I}$, i.e.,

$$\widehat{se_F(\widehat{I})} = se_{\widehat{F}}(I^{*b}) = \left[\sum_{b=1}^{B} (I^{*b} - \overline{I}^*)^2/(B-1) \right]^{1/2}. \tag{11}$$

Often, the α-th lower and upper percentile of the bootstrap distribution of the I^* (denoted by I^*_{lo} and I^*_{hi}) are used directly as the boundary values of $100 - 2\alpha$ per cent confidence interval for I. This, however, does not accord with the classical definition of a confidence interval[9]. A more appropriate confidence interval – known as the "Hall's percentile method" or "bias corrected confidence interval" – is the interval $\left(2\widehat{I} - I^*_{hi}\ ,\ 2\widehat{I} - I^*_{lo}\right)$, where I^*_{lo} and I^*_{hi} were defined above[10]. Obviously, the two methods produce the same interval if $\widehat{I}$ is the midpoint between I^*_{lo} and I^*_{hi}.[11]

Of greater interest is the assessment of the statistical significance of the change in inequality between two years[12]. Suppose we have two samples of sizes n_1 and n_2, drawn from completely unspecified probability distributions F_1 and F_2, and let $\widehat{\Delta I}$ denote the difference between the estimated inequality coefficients for the two samples, i.e. $\widehat{\Delta I} = \widehat{I}_1 - \widehat{I}_2$. The bootstrap procedure simulates the distribution of the difference of inequality estimates based on samples of n_1 and n_2 observations drawn from the empirical distributions $\widehat{F_1}$ and $\widehat{F_2}$. This is done by drawing bootstrap samples of sizes n_1 and n_2 from the empirical distributions

[9] See Hall (1994) or Efron and Tibshirani (1993).

[10] Note that although I^*_{lo}, $\widehat{I}$ and I^*_{hi} are all within the theoretical bounds of the inequality measure, $2\widehat{I} - I^*_{hi}$ and $2\widehat{I} - I^*_{lo}$ may not be. This can be remedied, however, by backing out the confidence interval for I from the bootstrap confidence interval for an unbounded one-to-one transformation of I. For further information, see Efron and Tibshirani (1993).

[11] See Hall (1994) for the use, performance and need of bias corrected intervals especially from skewed distributions.

[12] In the empirical application in section 4 we test whether inequality has significantly changed over time in Australia.

$\widehat{F_1}$ and $\widehat{F_2}$ respectively, and recording the difference between inequality estimates of the two samples. When this is repeated B times, it produces

$$\Delta I^{*b} = I_1^{*b} - I_2^{*b} \quad \text{for} \quad b = 1, ..., B. \tag{12}$$

Hall's percentile confidence interval for the difference in inequality between the two samples can be calculated from the bootstrap distribution using

$$\Pr(2\widehat{\Delta I} - \Delta I_{hi}^{*} \leq \Delta I \leq 2\widehat{\Delta I} - \Delta I_{lo}^{*}) = \frac{100 - 2\alpha}{100}, \tag{13}$$

where ΔI_{lo}^{*} and ΔI_{hi}^{*} are the α-th lower and upper percentile of the bootstrap distribution of the difference in inequality. If this confidence interval does not include zero, we can conclude that the change in inequality has been statistically significant.

(ii) Bootstrapping from samples collected through stratification

Biewen (2002) suggests the following method for bootstrapping when sample observations have unequal weights. Consider the sample empirical distribution $\widehat{F}$ as the distribution of (income,weight) pairs. Draw B bootstrap samples of n (income,weight) observations from $\widehat{F}$, i.e.,

$$\widehat{F} \longrightarrow \{(y_1^{*b}, w_1^{*b}), (y_2^{*b}, w_2^{*b}), ..., (y_n^{*b}, w_n^{*b})\} \qquad \text{for} \quad b = 1, 2, ..., B. \tag{14}$$

The weighted estimate of the inequality measure of interest is calculated for each sample, leading to a bootstrap distribution for the weighted estimator under $\widehat{F}$. The procedure of deriving estimated standard errors and confidence bands from the bootstrap distribution is as explained in the previous subsection. We have used Biewen's method to calculate all standard errors and confidence intervals estimated from the household expenditure survey data that are reported in the next section.

Although Biewen's method is "asymptotically" valid – i.e., when our sample covers the entire population $\left(\widehat{F} = F\right)$ and weights are all equal to 1, then this method produces the true sampling distribution of the inequality estimator based on samples of n observations – we think that it exaggerates the uncertainty of the estimates. This is because some of the information in the weights is disregarded in the resampling process. For example, weights are often assigned such that their sum equals to the estimate of *population* (as in the number of people living in a country). However, sum of the weights in the bootstrap samples can be wildly different from the *population*. As an alternative to Biewen's method, we considered that the empirical distribution $\widehat{F}$ places probability p_i on observation i, for $i = 1$ to n. We drew bootstrap samples from this distribution, and calculated the (unweighted) inequality measure for each of these samples. As expected, the confidence intervals derived this way were always tighter than the ones calculated from Biewen's method. However, since the difference was not large enough to affect our conclusions, we only report the confidence intervals calculated by the Biewen's method in the next section.

IV Empirical Application and Results

The empirical application is divided into two main parts. In the first part, section 4.1, we study income inequality within and between various geographical regions in Australia over the decade 1986 to 1996. In the second part, section 4.2, we study income inequality within and between various subgroups of the Australian population based on age, gender employment and occupation.

(i) Australian Inequality using Census Data

Using census data our analysis was naturally divided into two main sections, due to the nature of the data. We first examine the change in inequality over the decade of 1986 to 1996. Then due to the detailed coverage of the geographical regions in each the 1991 and 1996 census, we also study in more detail regional inequality in Australia.

The Data

The data available for this part of our application consists of three samples. They are a one percent sample for each of the 1986, 1991 and 1996 census. These are made available to by the Australian Bureau of Statistics (ABS) as Household Sample Files[13]. We employ in

[13] The 1986 Household Sample File is available as ABS catalogue no. 2196.0, the 1991 Household Sample File is available as ABS catalogue no. 2913.0, and the 1996 Household Sample File is available as ABS catalogue no. 2037.0.

our analysis annual income for individuals between the age of 18 and 64[14]. Unfortunately actual incomes are not reported as the income variable is categorized. The estimation of inequality measures from the categorical income variable for Australia has been studied by Chotikapanich and Griffiths (2000). These authors estimate the Gini coefficient once indirectly through estimating the parameters of the underlying density of income, and again directly by assuming that all individuals in a category earn the same income (equal to the midpoint of the category, or other plausible alternatives). Since the results produced by these two methods of estimation are quite similar[15], we use the second method and assign the midpoint income to all individual in the same category. After the appropriate cleaning and adjustment of the data, the 1986 sample has 85,471 individuals grouped into six geographical regions. The 1991 sample consists of 95,478 individuals divided in 20 geographical regions and the 1996 sample consists of 103,300 individuals grouped into 41 geographical regions. Due to the difference of the geographical regions in the three samples the 1991 and 1996 data were aggregated to geographically match the 1986 sample, as shown in the appendix, table 1.

Australian inequality 1986-1996: analysis and results

Table 1 presents the Gini coefficient (G) and the Theil measure (T) , (bootstrap standard errors in brackets), calculated from each of the 1986, 1991 and 1996 samples. The values of the Gini coefficient and the Theil indicate that income inequality in Australia is present to a significant degree[16]. Moreover there is a constant increase for both measures over the decade of 1986 to 1996. The decomposable nature of the Theil allows us to examine whether the inequality present in Australia is due to inequality within (WT) the regions, or between

[14] It should be noted that the income data on individuals in the 1986 census starts from the age of 20 due to the design of the age variable. Also for all three samples only individuals with some income were considered.

[15] This is our observation.

[16] In comparison the Gini coefficients published for the US by the US Census Bureau are 0.425 for 1986, 0.428 for 1991 and 0.455 and Theil's inequality measure is 0.31 for 1986, 0.313 for 1991 and 0.386 for 1996. It should be noted that the US coefficients are derived from household income data and not from personal income. For a general indication of the inequality present in Australia on an international scale see Gottschalk and Smeeding (1997).

(BT) the regions. Observing the between and within Theil decomposition, the inequality present in Australia is significantly due to inequality present within each region rather than between the regions.

	1986	**1991**	**1996**
G	0.3985	0.4091	0.4262
(se)	(0.0026)	(0.0027)	(0.0027)
T	0.2696	0.2816	0.3059
(se)	(0.0035)	(0.0037)	(0.0038)
BT	0.0009	0.0007	0.0007
(se)	(0.0007)	(0.0006)	(0.0006)
WT	0.2687	0.2809	0.3053
(se)	(0.0036)	(0.0037)	(0.0038)

Table 1: Gini Coefficient and Theil's Inequality measure with bootstrap standard errors.

In table 2 the pair wise difference between the Gini coefficients (ΔG) for the three samples are presented. Also presented are the confidence intervals of these differences and the percentage changes ($\%\Delta$) from the earlier, chronologically, to the later samples. The changes in the Gini coefficients are statistically significant (i.e. significantly different from zero). An important observation from this table is that the percentage increase in income inequality between 1991 and 1996 is approximately double the increase between 1986 and 1991.

		1991	**1996**
1986	ΔG	0.0105*	0.0277*
	(95% CI)	(0.008,0.012)	(0.026,0.029)
	$\%\Delta$	2.65%	6.94%
1991	ΔG		0.0171*
	(95% CI)		(0.016,0.018)
	$\%\Delta$		4.18%

Table 2: Confidence Intervals for the change in the Gini coefficient between 1986, 1991 and 1996.

From the above results we can clearly conclude that income inequality has significantly increased through the decade under consideration. An important question for Australia would be to ask what this inequality is attributed to. For example is this income inequality increase due to state policies (i.e. is the increase attributed to the between regions increase in inequality) or is it an all Australian phenomenon (i.e. within the regions increase in income inequality). Considering the changes in the Theil over the decade, table 3 indicates that the increase in inequality is only attributed to within region inequality increase. In fact there seems to be a decrease in income inequality between regions although it is statistically insignificant.

		1991	1996
1986	ΔT 95% CI	0.0119* (0.002,0.022)	0.0364* (0.026,0.046)
	ΔBT 95% CI	−0.0002 (−0.002,0.002)	−0.0002 (−0.002,0.002)
	ΔWT 95% CI	0.0122* (0.001,0.022)	0.0366* (0.026,0.047)
1991	ΔT 95% CI		0.0244* (0.014,0.035)
	ΔBT 95% CI		−0.00005 (−0.001,0.002)
	ΔWT 95% CI		0.0244* (0.014,0.035)

Table 3: Confidence Intervals for the change in the Theil's Entropy measure and its decompositions between 1986, 1991 and 1996.

Australian Regional Inequality 1991 and 1996: Analysis and Results

As it was concluded in the previous section, the income inequality present in Australia, is due to inequality within the geographical regions around Australia rather than between them. This section is an extensive study of income inequality within geographical regions via the Gini coefficient. In contrast to the 1986 census, the 1991 and 1996 data is divided into smaller geographical regions within the states and territories. The 1991 data is divided into 20 regions, and the 1996 data divided into 41 regions. The 1996 data was aggregated into the 20 regions (presented in the appendix, table 2) to match the 1991 census, as the 41 regions where subdivisions of the 1991 regions. It should be mentioned that this study is more extensive for the states of New South Wales, Victoria and Queensland as these three states are extensively subdivided in both samples.

Table 4 presents the Gini coefficients of each region for each of the 1991 and 1996 census. The differences between these is presented in column 5 with the 95% confidence interval bellow each change. The regions have been ranked within each state, based on the magnitude of the change in income inequality from 1991 to 1996. These results clearly indicate significant increases in inequality for each of the regions under consideration, from 1991 to 1996, or at least any decrease seems to be insignificant. Within each of the first three states, New South Wales, Victoria and Queensland, inequality seems to have increased more around the metropolitan areas. This increase diminishes approaching rural areas[17]. For example the regions 8 to 12 are part of Victoria. Regions 8,9 and 10 are metropolitan areas around the city of Melbourne and regions 11 and 12 are part of rural Victoria.

State	**Region**	**G** (96)	**G** (91)	**ΔG** (96 − 91) (95% CI)
NSW	2	0.4166 (0.009)	0.3922 (0.009)	0.0244* (0.0205,0.0283)
	4	0.4307 (0.008)	0.4101 (0.008)	0.0206* (0.0174,0.0283)
	3	0.4090 (0.009)	0.3898 (0.009)	0.0192* (0.015,0.0236)
	1	0.4169 (0.008)	0.4 (0.008)	0.0169* (0.0131,0.0206)
	5	0.4319 (0.008)	0.4205 (0.008)	0.0114* (0.0071,0.016)
	7	0.4105 (0.009)	0.4026 (0.009)	0.0078* (0.0035,0.0124)
	6	0.4144 (0.009)	0.4143 (0.009)	0.0 (−0.005,0.0056)
VIC	8	0.4267 (0.009)	0.3897 (0.009)	0.037* (0.0327,0.0412)
	9	0.4427 (0.009)	0.4202 (0.008)	0.0225* (0.019,0.0259)
	10	0.4117 (0.009)	0.3920 (0.009)	0.0198* (0.0157,0.0239)
	11	0.4143 (0.009)	0.3997 (0.009)	0.0146* (0.0101,0.0196)
	12	0.4117 (0.009)	0.4143 (0.009)	−0.003 (−0.007,0.0023)
QLD	15	0.4286 (0.009)	0.4109 (0.009)	0.0177* (0.0128,0.023)
	14	0.4186 (0.009)	0.4055 (0.009)	0.0132* (0.0084,0.018)
	13	0.4142 (0.009)	0.4025 (0.009)	0.0117* (0.0078,0.0159)
	16	0.4180 (0.009)	0.4212 (0.009)	−0.003 (−0.008,0.0026)
Other	17	0.4092 (0.009)	0.3972 (0.008)	0.012* (0.0078,0.0163)
	18	0.4288 (0.009)	0.4138 (0.009)	0.015* (0.0106,0.0195)
	19	0.4322 (0.008)	0.4152 (0.008)	0.017* (0.0124,0.0214)
	20	0.4228 (0.009)	0.4085 (0.008)	0.0143* (0.0122,0.0182)

Table 4: Regional changes in Gini coefficient around Australia

(ii) Australian Inequality using Household Expenditure Survey datasets

In studying inequality from Household expenditure data we have divided our analysis

[17] For more details on which areas are considered metropolitan and which rural see the ABS Australian Standard Geographic Classification catalogue number 1216.0.

into two main parts. We first study income inequality for all individuals over the age of 15 and then restrict our sample to employed and self-employed individuals between the ages of 15 and 64.

The Data

The data employed is made available by the Australian Bureau of Statistics from the Household Expenditure Surveys[18] for 1993/94 and 1998/99. The greatest advantage of the household expenditure survey data sets, in comparison to the census data, is that actual weekly incomes for individuals are reported. As mentioned in the introduction these are stratified data sets and the sampling weights are reported. We make use of the sampling weights to calculate weighted inequality measures and bootstrap standard errors and confidence intervals, as described in the bootstrap methodology in section *III(ii)*.

Australian inequality 1993-1998: All individuals above the age of 15 years

After adjusting and cleaning the data sets, we have 13,964 individuals for the 1998/99 sample and 17,271 individuals for the 1993/94 sample. Table 5 presents the calculated weighted Gini (G_w), weighted Theil (T_w) and the changes in the inequality measures from the 93/94 to 98/99. Comparing these results to the earlier inequality measures from the census data, these measures, as was expected, are relatively higher than the ones presented in table 1. Besides the higher variability present in the weekly income in comparison to annual income[19], the income distribution of the census data is truncated at the midpoint of the first and the last of the income categories. This further reduces the variability in the income distribution. Interestingly in contrast to the significant increases of both Gini and Theil over the decade of 1986 to 1996, between 1993/94 and 1998/99 only the increase in the Gini coefficient is found to be statistically significant.

[18] The 1993/94 Household Expenditure Survey is available as ABS Catalogue No. 6544.0.15.001. The 1998/99 Household Expenditure Survey is available as ABS Catalogue No. 6544.0.30.001.

[19] See footnote 3 page 1.

	1993/94 (*se*)	1998/99 (*se*)	Change in Inequality (95% *CI*)
G_w	0.4815 (0.003)	0.4931 (0.0029)	0.01154* (0.0033,0.0193)
T_w	0.4181 (0.0081)	0.4303 (0.0066)	0.0122 (−0.0058,0.0325)

Table 5: Weighted Gini and Theil and the change in the measures from 93/94 and 98/99.

To further investigate the inequality within Australia we have made use of the Theil decomposition. We investigate possible sources of inequality and also how inequality has changed over the five year period the amongst various subgroups. The first decomposition we perform in that between males and females presented in table 6. The Theil decomposition, for both years, indicates that income inequality is vastly attributed to inequality within each of the subgroups, rather than inequality between males and females. Also there seems to be no significant increase in inequality over the five years either within or between these groups.

	1993/94 (*se*)	1998/99 (*se*)	Change in Inequality (95% *CI*)
T_w	0.4181 (0.0081)	0.4303 (0.0066)	0.0122 (−0.0086,0.0325)
BT_w	0.0269 (0.0018)	0.0302 (0.002)	0.0032 (−0.0021,0.0088)
WT_w	0.3912 (0.0076)	0.4001 (0.0061)	0.0089 (−0.01,0.0274)

Table 6: Theil decomposition between males and females

The second set of subgroups we investigate are the age-groups. There are thirteen age-groups as shown in table 8. The Theil presented in table 7 indicates that inequality is attributed to within age-groups inequality and there seems to be no significant increase in the inequality over the five year period either between or within the age-groups. The Theil measures for each of the age-groups are presented in table 8.

	1993/94 (*se*)	1998/99 (*se*)	Change in Inequality (95% *CI*)
T_w	0.4181 (0.0081)	0.4303 (0.0066)	0.0122 (−0.0086,0.0325)
BT_w	0.075 (0.0026)	0.0796 (0.0027)	0.0045 (−0.0028,0.0116)
WT_w	0.3432 (0.0075)	0.3512 (0.0061)	0.0077 (−0.0124,0.0264)

Table 7: Theil decomposition by age groups

From the Theil measures of each age-group, it is interesting to note that the highest

Theil (i.e. most unequal group) recorded in this paper, is for individuals between the ages of fifteen and nineteen years. A general observation from table 8 is that inequality has generally decreased between the ages of fifteen and forty four, with the last of these age-groups (forty to forty four) experiencing a statistically significant decrease. Also inequality seems to have generally increased for individuals above the age of forty four with the subgroups of individuals above the age of seventy five experiencing a statistically significant increase.

Groups	**Ages**	$\mathbf{T}_w$ (93/94)	$\mathbf{T}_w$ (98/99)	ΔT_w (98/99 − 93/94) (95% CI)
1	15 − 19	0.7519	0.7407	−0.0111 (−0.0832,0.0619)
2	20 − 24	0.2667	0.2709	0.0042 (−0.0553,0.0794)
3	25 − 29	0.2429	0.2374	−0.0055 (−0.0362,0.0259)
4	30 − 34	0.3195	0.311	−0.0085 (−0.0476,0.0334)
5	35 − 39	0.3602	0.3481	−0.0121 (−0.0709,0.0542)
6	40 − 44	0.3889	0.3273	−0.0616⋆ (−0.1208,−0.0024)
7	45 − 49	0.3468	0.3959	0.0491 (−0.0087,0.1014)
8	50 − 54	0.4304	0.4566	0.0261 (−0.0501,0.1028)
9	55 − 59	0.4727	0.4749	0.0023 (−0.0686,0.0782)
10	60 − 64	0.3734	0.3703	−0.0031 (−0.0781,0.0713)
11	65 − 69	0.2346	0.2759	0.0413 (−0.0304,0.1129)
12	70 − 74	0.2124	0.3158	0.1035 (−0.016,0.2143)
13	75+	0.1481	0.2834	0.1343⋆ (0.066,0.1962)

Table 8: Inequality within age groups

Tables 9 and 10 present the analysis, of inequality being conditional on the five categories of the employment status of individuals. Table 9 indicates that inequality is attributed to both inequality within the subgroups (shown in table 10) and between them. Also there is a statistically significant increase in inequality from 93/94 to 98/99 in the inequality between these subgroups.

	1993/94 (*se*)	1998/99 (*se*)	Change in Inequality (95% *CI*)
T_w	0.4181 (0.0081)	0.4303 (0.0066)	0.0122 (−0.0086,0.0325)
BT_w	0.1835 (0.0037)	0.2079 (0.0039)	0.0245* (0.014,0.0349)
WT_w	0.2346 (0.0072)	0.2223 (0.0052)	−0.0123 (−0.0301,0.0038)

Table 9: Theil decomposition by employment status

Further analysing subgroups of individuals based on their employment status there seems to be no significant increase in the inequality within any of these subgroups and the most unequal subgroup is that of self employed individuals.

Groups	Employment Status	$\mathbf{T}_w$ (93/94)	$\mathbf{T}_w$ (98/99)	ΔT_w (98/99 – 93/94) (95% *CI*)
1	Full Time	0.1384	0.1409	0.0025 (−0.0156,0.021)
2	Part Time	0.2982	0.2491	−0.0491 (−0.0954,0.001)
3	Self Employed	0.5272	0.5432	0.016 (−0.0749,0.1068)
4	Unemployed	0.3762	0.4116	0.0354 (−0.0306,0.1026)
5	Not in workforce	0.4413	0.4141	−0.0272 (−0.073,0.0227)

Table 10: Inequality within employment status groups

Australian inequality 1993-1998: Employed and self-employed individuals between the ages of 15 and 64 years

The interesting results of inequality among all individuals conditional on their employment status led us to the study of inequality among individuals participating in the labor force. 7,931 individuals in the 1993/94 sample and 6,375 individuals in the 1998/99 sample were employed full time or self employed. The Theil presented in table 11 shows that inequality conditional on the occupation of individuals is primarily attributed to inequality within each of the occupation groups rather than between them.

	1993/94 (*se*)	1998/99 (*se*)	Change in Inequality (95% *CI*)
T_w	0.2032 (0.0081)	0.2062 (0.0064)	0.0029 (−0.0183,0.0234)
BT_w	0.0211 (0.0017)	0.0225 (0.0019)	0.0013 (−0.0039,0.0066)
WT_w	0.1821 (0.0076)	0.1837 (0.0056)	0.0016 (−0.0177,0.0196)

Table 11: Theil decomposition by occupation

Table 12 presents the Theil within each of the occupation groups. The highest Theil recorded is for managers and administrators who experience a statistically significant decrease in inequality from 93/94 to 98/99. Statistically significant increases in inequality have been observed for associate professionals, labourers and related workers.

Groups	**Occupation**	$\mathbf{T}_w$ (93/94)	$\mathbf{T}_w$ (98/99)	ΔT_w (99 − 94) (95% CI)
1	Managers and administrators	0.3765	0.2972	−0.0792* (−0.1475,−0.007)
2	Professionals	0.1741	0.176	0.0019 (−0.0365,0.0432)
3	Associate Professionals	0.068	0.1928	0.1248* (0.0933,0.1582)
4	Trades persons	0.1595	0.191	0.0315 (−0.003,0.0673)
5	Production workers and drivers	0.1660	0.163	−0.003 (−0.0668,0.0778)
6	Clerks, sales and service	0.127	0.1278	0.0007 (−0.0204,0.0214)
7	Labourers and related workers	0.1038	0.1619	0.0581* (0.0244,0.0927)

Table 12: Inequality within occupation groups[20]

V Conclusion

In this paper we have studied the changes in income inequality in Australia between 1986 and 1999 via the Gini coefficient and Theil's Inequality measure. We have demonstrated the bootstrap procedure for independent and stratified samples. Using these bootstrap methods we have estimated standard errors and confidence intervals and we have established the statistical significance of changes in income inequality over time and within various population groups. Our main findings based on census data are that the distribution of income is significantly more unequal in 1996 relative to 1986 and the increase in inequality has been due to increase in inequality within geographical regions rather than between them. We also found that the increase in inequality has been more pronounced around major metropolitan areas. Our analysis based on HES data suggests that the upward trend in inequality may have leveled out in the late 90's. It would be interesting to know if the

[20] For further details on the classification of the occupation groups see the Australian Standard Classification of Occupations (2nd edition) avalaible on the ABS website http://www.abs.gov.au

newly released 2001 Census data corroborates this evidence. We find that there has been a general decline in inequality among younger people (less than 45), but an increase in inequality among older people, with quite significant increase in income inequality among the 75+ group. We think that the sharp increase in house prices which on the one hand, has contributed to older generations being *on average* richer than a decade ago, may have also created a large divide between owners and non-owners of property in this age group.

References

Atkinson A.B., (1983), *The Economics of Inequality*, Clarendon Press Oxford, New York, 2nd Edition.

Bernstein, J., Mishel, L. and Brocht, C. (2001), 'Any Way You Cut It Income Inequality on the Rise Regardless of How It's Measured', *Economic Policy Institute*, Briefing Paper, Washington DC.

Biewen, M. (2002), 'Bootstrap Inference for Inequality, Mobility and Poverty Measurement', *Journal of Econometrics* **108**, 317-342.

Borland, J. (1999), 'Earnings Inequality in Australia: Changes, Causes and Consequences', *Economic Record* **75**, 177-202.

Chotikapanich, D. and Griffiths, W.E. (2000), 'Posterior Distribution for the Gini Coefficient using Grouped Data', *Australian and New Zealand Journal of Statistics* **4**, 901-910.

Coulter, F.A.E., Cowell, F.A. and Jenkins, S.P. (1992), 'Equivalence Scale Relativities and the Extent of Inequality and Poverty', *The Economic Journal* **102**,1067-1082.

Cowell, F. (1995), *Measuring Inequality*, LSE Handbook in Economics Series, 2nd edn, Prentice Hall, London.

Efron, B. (1979), 'Bootstrap Methods another look at the Jacknife', *Annals of Statistics* **7**, 1-26.

—– and Tibshirani, R. (1993), *Introduction to the Bootstrap*, New York, Springer-Vergal.

Gini, C. (1912), *Variabilità emutabilità*, Bologna.

Gottschalk, P. (1997), 'Inequality, Income Growth and Mobility: The Basic Facts', *Journal of Economic Perspectives* **11**, 21-40.

—– and T.M. Smeeding (1997), 'Cross-National Comparisons of Earnings and Income Inequality', *Journal of Economic Literature* **35**, 633-687.

Hall, P. (1994), 'Methodology and Theory for the Bootstrap', *Handbook of Applied Econometrics* **4**, eds Engle, R. and McFadden, D. Amsterdam: North Holland.

Harding, A. (1994), 'Income Inequality in Australia from 1982 to 1993: An Assessment of the Impact of Family, Demographic and Labour Force Change', *NATSEM Discussion Paper,* No **4**.

Karoly, L.A. and Burtless, G. (1995), 'Demographic Change, Rising Earnings Inequality and the Distribution of Personal Well-Being 1959-1989', *Demography*, **32**, 379-405.

Mills, J.A. and Zandvakili, S. (1997), 'Statistical Inference via Bootstrapping for Measures of Inequality', *Journal of Applied Econometrics*, **12**, 133-150.

Sen, A. (1993), *On economic Inequality*, Clarendon Press, Oxford.

Theil, H. (1967), *Economics and Information Theory*, North-Holland, Amsterdam.

APPENDIX

REGION	AREA
1	New South Wales
2	Victoria
3	Queensland
4	Western Australia
5	South Australia
6	Australian Capital Territory Northern Territory Tasmania

Table 1: The six regions considered in section 4.1.2 consisting of the six states and the two territories.

STATE	REGION	Area within State included in region
NSW	1	Inner Sydney/ Inner Sydney/ Eastern Suburbs
	2	Sutherland/ Canterbury/ Bankstown/ Liverpool
	3	OuterSW/W Sydney/Inner and Central W Sydney
	4	Lower N Sydney/Northern Beaches
	5	Hunter/ Illawarra/ South Eastern
	6	Northern NSW
	7	Central West/ Murray/ Murrumbidgee
VIC	8	OuterW /NW /NE /Inner Melbourne
	9	Inner East/ S Melbourne
	10	OuterE/ SE/ Melb and Mornington Peninsula
	11	Western Victoria (Loddon/Mallee/Wimmera)
	12	EastVictoria (Gippsland/Ovens-Murray/Goulburn)
QLD	13	Brisbane
	14	Moreton
	15	North QLD
	16	Remainder QLD
Other	17	Adelaide
	18	Perth
	19	Remainder SA,WA
	20	Tasmania/ACT/NorthernTerritory

Table 2: The 20 regions considered in section 4.1.3 as provided by the Australian Bureau of Statistics.