Understanding Changes in Māori Incomes and Income Inequality 1997-2003

Sylvia Dixon and David C Maré

Motu Working Paper 04-12
Motu Economic and Public Policy Research

December 2004

Author contact details
Sylvia Dixon
Research Analyst
Motu Economic and Public Policy Research
Level 1, 93 Cuba Street
PO Box 24390
Wellington
Email: sylvia.dixon@motu.org.nz

Dave Maré
Senior Fellow
Motu Economic and Public Policy Research
Level 1, 93 Cuba Street
PO Box 24390
Wellington
Email: dave.mare@motu.org.nz

Acknowledgements
This research was funded by the Foundation for Research, Science and Technology, through its support for Motu's "Understanding Adjustment and Inequalities" research programme. The financial support of the Foundation is gratefully acknowledged.

Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the authors, not Statistics New Zealand.

The authors would like to thank Robert Templeton and Jonathan Khoo for their valuable assistance with the calculation of sampling errors. They also thank Steven Stillman, participants at the 2004 NZ Association of Economists conference, and participants at the 2004 New Directions: New Settlers: New Challenges workshop for their helpful comments on earlier drafts of this paper.

Motu Economic and Public Policy Research
PO Box 24390
Wellington
New Zealand

Email	info@motu.org.nz
Telephone	+64-4-939-4250
Website	www.motu.org.nz

 ISSN 1176-2667.

Abstract

This paper reports findings from a study of changes in Māori income levels and income dispersion between 1997 and 2003. Data from Statistics New Zealand's Income Survey are used to describe and evaluate the main changes in the Māori income distribution in this period, which was marked by substantial increases in employment rates and improvements in the skill levels of working-aged Māori. A parallel analysis of the main changes in the European/Pākehā income distribution is provided for comparative purposes.

The results show significant reductions in the proportion of Māori with no weekly income in the reference week, or incomes of $150–200 a week, and significant increases in the proportion with incomes above the peak income level of approximately $550 per week. Income inequality within the total working-aged Māori population declined, while income inequality among employed Māori was stable. An analysis of some of the key factors contributing to change in the income distribution suggests that the transition of many Māori into employment during this period was the single most important driver of change.

JEL classification
D31 (Personal income and wealth distribution); J15 (Economics of minorities and races).

Keywords
Individual income distribution; Inequality; Māori; Kernel density estimation.

Contents

Table of figures

Tables

1 Introduction

Strong economic and employment growth during the late 1990s and early years of the 21st century was accompanied by substantial increases in levels of economic activity among Māori. Movements in the published indicators of aggregate Māori employment rates and average incomes offer evidence of changes on a scale that would be expected to lead to some real improvements in the material well-being of Māori. The published indicators, however, convey little information about the distribution of those improvements *within* the Māori population.

This paper examines the changes that have occurred in the Māori income distribution at individual level, during the past six years, using data from Statistics New Zealand's Income Survey (IS). The income distribution is analysed because income is both an important source of material well-being and strongly correlated with other dimensions of living standards, such as quality of housing.

To date, research on the incomes of Māori has focused almost entirely on changes in *average* Māori incomes or on the Māori/non-Māori income gap. National averages have the potential to conceal wide variations in outcomes between individuals or sub-groups. The distributional focus of this paper is prompted by that gap in existing knowledge.

The Working Paper has three main objectives. The first is simply to accurately describe and evaluate the main changes in the Māori income distribution between 1997 and 2003. The Income Survey is a relatively new survey (beginning in 1997), and to our knowledge we are the first to attempt to use this data source for this purpose. We use descriptive statistics and kernel density estimation methods to identify the changes that occurred in different regions of the income distribution and the overall impact on income dispersion.

A second aim of the paper is to evaluate the capacity of the Income Survey to provide accurate and reliable information on changes in Māori incomes over a period such as five years—taking into account the fact that the sample of Māori respondents in the Income Survey is relatively small, and the Income

Survey has complex design features that reduce the accuracy of its estimates, compared with a simple random sample. We calculate sampling errors on key estimates of income change in a manner that takes into account the main survey design features, and interpret our results with reference to that sampling error information.

A third aim of the paper is to explore the most likely causes of the income distribution changes. Our approach to this problem is to estimate a series of 'counterfactual' income distributions that model the effects of different factors on the income distribution for working-aged Māori, including changes in the age structure, changes in educational levels, shifts in employment rates, and changes in labour market activity patterns. This analysis offers some clues as to which of the potential sources of change are likely to have had most impact on the level and shape of the Māori income distribution.

The paper begins with a brief review of previous research on the subject. The data source and methods are outlined in Section 3. Section 4 gives background information on major changes in the labour market activity patterns and demographic characteristics of Māori in this period. The core results of the paper are presented in Section 5. Section 6 reflects on the findings and concludes.

We find evidence of significant reductions in the proportion of working-aged Māori with zero incomes or gross weekly incomes of around $150–200 a week (measured in June 2000 dollar values). There were corresponding increases in the proportion of Māori with gross weekly incomes above $500 a week, particularly in the $500–600, $700–870, $1,100–1,500 and $1,900–2,500 ranges. These changes are consistent with the impact one would expect from a large increase in the aggregate Māori employment rate. Focusing on the incomes of *employed* Māori, there are signs of real income growth at all levels of income. While the precise changes are more difficult to identify with confidence, there were some significant increases in the proportion of employed Māori at income levels between $1,200 and $2,200 per week.

The distribution of weekly income across working-aged Māori became less dispersed (more equal) in this period. This was due, in large part, to the transition of many Māori who were previously not working into employment. The dispersion of income among employed Māori did not change significantly.

Analysis of the sources of change in the income distribution of all working-aged Māori suggests that the increase between 1997/98 and 2002/03 in Māori full-time and part-time employment rates was likely to have been the single most important driver of change, operating over all regions of the income distribution. Improvements in the educational qualifications of Māori and occupational change may also have had a significant impact, particularly in the upper half of the income distribution.

Shifts in the Pākehā/European income distribution in this period were broadly similar in size and direction to the changes in the Māori income distribution, suggesting some common drivers of change.

2 Previous research

The distribution of income within the Māori population is a relatively neglected subject. Past researchers have generally focused on changes in average Māori incomes or the Māori/non-Māori income gap, not on distributional issues.

Richard Benton and colleagues (2003) have investigated disparities within the Māori population living in the Greater Auckland region, using a combination of ethnographic research and analysis of existing statistical data (drawn mostly from the 1996 Census). They report that there is a great deal of variation in individual-level annual income levels *within* each geographic locality studied; between localities; between iwi; and within iwi (Benton, 2003, pp.36–41). The extent of income dispersion is briefly assessed in the research through a comparison of Gini coefficients for each sub-group of Māori included in the study. The analysis is cross-sectional in nature and does not consider changes in Māori income levels or dispersion since 1996.

Chatterjee and Podder (2003) present information on changes in household-level inequality within each of New Zealand's major ethnic groups, calculating the Gini coefficient of the concentration of gross annual household income, and using this as their sole measure of inequality. The data are from the

Household Economic Survey. Their estimates suggest that the inequality of household incomes for the Māori and Pacific peoples ethnic groups (combined) declined during the period 1984–1996, before increasing again (but to a lower level) in 1998.

Maani (2000; 2002) has examined changes in the average annual incomes of Māori and Europeans between 1986 and 1996, using micro-data from the population census. She analyses the rate of growth in Māori incomes from the mid-1980s to the mid-1990s, and the influence of factors such as education, age and residential patterns on Māori incomes at the mean. The focus is on explaining Māori–Pākehā income differences, and in particular the contribution of educational disparities to those income differentials by ethnic group. The study does not consider the distribution of income within the Māori population, however.

3 Data and methods

3.1 Data source, sample and variable definitions

Statistics New Zealand (SNZ) gathers three data sources that could potentially be used to study the Māori income distribution: the Household Economic Survey, the Income Survey, and the Population Census. Each has some advantages and disadvantages. The Income Survey collects information on individuals' actual pre-tax weekly income in the reference week for the survey, and uses a series of detailed questions on each component of income in an effort to ensure that the final estimate of total income (from all sources) is as accurate as possible. It provides a much larger sample of Māori respondents than does the HES, and an unbroken annual series of results. The main disadvantage of the Income Survey as a data source is the relatively short time period covered (1997–2004). The Census, while offering a much larger sample of Māori and covering a longer time period, has the disadvantage of measuring income rather crudely, in broad bands.

The Income Survey is carried out by Statistics New Zealand each June quarter, as a supplement to the Household Labour Force Survey (HLFS). Taken together, the two surveys collect data on household structure, the sociodemographic characteristics of household members, labour force activity in the reference week, and recent incomes. The HLFS has a sample size of approximately 15,000 households and 28,000 adults. About 85% of these respondents also complete the Income Survey. Sampling weights are calculated by SNZ, and these are used in this analysis.

The population of study is restricted to adults aged 20–59 years. This age range approximately captures the age groups in which the majority of members (more than half) are engaged in the labour market. The 'Māori' sample includes all those who specified 'Māori' as one of their ethnic identities. This is the most inclusive and commonly used definition of 'Māori'.[1] For comparative purposes, we also report results from a parallel analysis of the incomes of the European ethnic group. The European sample comprises all those who specified 'Pākehā', 'European', or any specific European ethnic group, and did not affiliate with any non-European ethnic group. It includes both New Zealand-born and overseas-born Europeans.[2]

The small number of Māori who are included in the samples of most general population surveys represents a challenge for empirical research on this ethnic group. Table 1 gives information on sample sizes that were available for this study. There are about 1,100 Māori men aged 20–59 in each annual IS sample, and about 1,400 women of this age group. In response to the fact that these sample sizes are relatively small, the 1997 and 1998, and the 2002 and 2003 samples were pooled in all analyses undertaken in this study to reduce the effects of sampling variability on the estimates obtained, and reduce sampling errors. 1997/98 represents the starting period for the analysis of change and 2002/03 the end period, giving an average gap of five years.

Around 15% of Income Survey responses are imputed by SNZ, because the respondent was not available to answer the questions in person. The imputation rate is somewhat higher for Māori than for other ethnic groups (18% on average during the study period). Imputed responses were used in the

[1] It matches the ethnic category 'Māori' at the highest (1-digit) level of the official classification.
[2] This corresponds to the ethnic category 'European' at the highest (1-digit) level of the official classification.

calculation of the main results reported in this paper, but not in the calculation of sampling errors. Table 2 gives information on imputation rates for various sub-populations, and the sample sizes that remained once the imputed records were removed.

The income measure used in this study is actual gross weekly income from all sources, as received in the reference week (the week before the interview). This includes actual gross weekly earnings, income from self-employment, income from government benefits, national and private superannuation, student allowances, and earnings-related compensation received from the Accident Compensation Corporation (ACC).[3] Income from investments is not included, because it was not measured systematically in the IS until 2002.[4] The income variable was converted into June 2000 dollar values for this analysis, using the Consumers Price Index (CPI).

3.2 Estimation methods

Kernel density estimation methods are used in this paper to estimate and visually depict the shape of the income distribution in 1997/98 and 2002/03. Kernel density estimation provides a means to identify the exact location in the income distribution where any changes have occurred. The method is briefly outlined in Appendix B. We also calculate a range of summary statistics of income level and dispersion at different points in the distribution, to better quantify the changes.

Standard errors and confidence intervals are calculated on all income change estimates to identify which changes are statistically significant. This is done using a modified bootstrap method that takes into account the effects of the key design features of the Income Survey that affect the variability of estimates.

[3] This differs slightly from the published measure of gross weekly income in using actual rather than usual earnings. This is consistent with the basis on which transfer income is measured (in terms of actual receipts).
[4] The investment income figures that have been collected since 2002 and published by Statistics New Zealand show that investment income is in fact a very small component of Māori weekly incomes.

Firstly, as in any sample survey, each observation in the Income Survey is differently weighted to reflect differences in the probability of selection. These weights, which are used to calculate accurate point estimates, also affect the size of sampling errors. Secondly, the IS sample design involves geographical clustering: respondents are drawn from particular strata and primary sampling units, not from everywhere in the country. Clustering increases sampling variability. Thirdly, the responses of about 15% of people in the sample are imputed by Statistics New Zealand because the individuals concerned could not be contacted to respond in person to the survey (see Appendix A). While imputation improves the quality of point estimates it does not increase the effective sample size for purposes of calculating standard errors, and therefore an adjustment must be made when standard errors are calculated. Estimates of the 'true' standard errors need to take account of each of these design features.

The bootstrap error estimation method used in this research uses the non-imputed records only, survey weights that have been adjusted to compensate for the exclusion of the imputed records, and information on the clustering of the sample.[5] The method is described in Appendix A. Results obtained using this method are similar to those obtained using the modified jackknife estimation method that is recommended by Statistics New Zealand. The sampling errors that were estimated using the final method were typically around 1.5 times the size of those obtained using the standard statistical software and sampling weights only (i.e. with no adjustments for sample clustering or imputation).

4 Background

4.1 The labour market and demographic context

Māori economic activity levels were particularly severely affected by the recession and economic reforms of the late 1980s and early 1990s. The aggregate Māori employment rate fell steeply between 1986 and 2002. This is illustrated in aggregate employment rate data shown in Figure 1, sourced from the published HLFS. From 1993 onwards, employment growth resumed and by 2003

[5] The modified weights and clustering information were supplied by Statistics New Zealand.

the aggregate Māori employment rate had returned to 1986 levels. It rose by 8 percentage points in the period of study (1997–2003) alone.

These large changes in employment rates could be expected to have significant flow-on impacts for Māori income levels and distribution—and indeed this study suggests that was the case. It is important to view the evidence analysed in this study within a longer-term context, however. The substantial improvements in employment and incomes that were recorded during the 1997–2003 period were in part a reversal of ground that was 'lost' earlier on.

Table 3 gives information on the changing employment patterns of the study population of working-aged Māori. It tabulates the employment rates and full-time employment rates of working-aged Māori in the base period (1997/98) and end period (2002/03). The total employment rate rose by 9.4 percentage points, and the full-time employment rate by 8.6 percentage points.

During the same period, there were some notable shifts in the income-related characteristics of this population (summarised in Table 4). It became relatively older, by 1.2 years on average. The educational profile of the population was lifted as an increasing proportion of Māori acquired post-school qualifications at both degree and sub-degree level. Whereas in 1997/98 36.6% of working-aged Māori held a post-school qualification, by 2002/03 this proportion had risen to 44.5%.

The occupational composition of employed Māori also shifted away from manual jobs in this period. The proportion of Māori who were working in clerical or in managerial, professional, or technical occupations expanded. At the same time the proportions working in trade occupations, in metal and machinery operative and processing jobs, and in elementary jobs declined.[6] There was little change in average weekly hours or the distribution of hours worked.

[6] The New Zealand Institute of Economic Research (NZIER) (2003) provides further information on changes in the industrial and sectoral composition of Māori employment.

4.2 Average real income increases

Table 5 shows the total increases that were recorded in the average inflation-adjusted incomes of our study sample. On average, working-aged Māori experienced a total of 15.6% growth in their real weekly incomes in this period, a little higher than the increase of 13.0% recorded for Pākehā/Europeans. The real income growth experienced by *employed* working-aged Māori was lower, at 8.2%. The difference between the two growth rates reflects the importance of movement into jobs as a factor raising average incomes for the total population.

During the same period, the proportion of Māori who reported zero weekly incomes in the reference week[7] declined from 10.2 to 8.1 percentage points. The proportion reporting receipt of benefit income from the Work and Income Service in the reference week declined from around 35% to 28%. There was also a small reduction in the average weekly incomes of people who were not in any form of employment, probably because of a change in the composition of this group towards a greater share of people who were not drawing income support benefits.

Māori women experienced faster real income growth in this period than Māori men, reducing the gender income differences. This was linked to a larger increase in the employment rate of Māori women.

5 Analysis of changes in the Māori income distribution

5.1 Kernel density estimates

We begin with a visual examination of the change in the shape of Māori income distribution, considering all Māori first and then each gender separately. A kernel density estimator was used to estimate the density of the income distribution at each level of real log weekly income.

[7] This is the week preceding the interview.

Specifically, the kernel density estimate $\hat{f}_h$ of an univariate distribution *f*, based on a random sample w_1, w_n with weights θ_1 , $\theta_{n,,}$ $\sum \theta_i = 1$, is

$$\hat{f}_h(w) = \sum_{i=1}^{n} \frac{\theta_i}{h} K\left[\frac{w - w_i}{h}\right]$$

where *h* is the bandwidth and *K*(.) is the kernel function. After presenting the core results, we discuss the statistical significance of the changes observed.

The income distributions for all working-aged Māori in 1997/98 and 2002/03 are shown in Figure 1. These can be interpreted in roughly the same way as one would read a histogram of relative frequencies. The x-axis represents weekly income, measured in June 2000 dollars and shown on a log scale. The y-axis represents the density of the income distribution. The labels on the x-axis have been converted from log to dollar values to aid interpretation. The total area under the income distribution curve is the same in each time period. Vertical differences in the height of the income distribution curve at any particular level of income represent increases/decreases in the proportion of the population at that point.

Note that in this part of the analysis, we have 'censored' the real weekly income data above 8.7 log points (approximately $6,000 in June 2000 dollar values) and below 2.3 log points (approximately $10). This is to limit the length of the tails of the distribution, and make it easier to plot the portion of the income distribution where the vast majority of people are located. Weekly income values above 8.7 were changed to 8.7, and values below 2.3 (of which the vast majority represent people reporting zero income in the reference week) were changed to 2.3. This change does not apply to or affect the analysis of income levels and dispersion in the rest of the paper.

Several features of change stand out in Figure 2. Firstly, there is a 'spike' at the beginning of the distribution, representing people who have no income in the reference week. The size of this spike declined between the base and end periods. Secondly, there was a hollowing out of the density at relatively low levels of income, between about $60 a week and $250 a week. A pronounced peak in the distribution, located between $150 and $200 a week, was much reduced in size. The central peak of the income distribution moved upwards and to the right, consistent with the growth in average real incomes described earlier. Finally, there was an increase in the proportion of people located at all mid-to-high levels of income, from about $650 per week upwards.

The decline in the size of the spike at zero incomes, and the decline in the proportion of Māori with incomes between $60 and $250 per week, could be explained, in part at least, by a transition of people who were previously out of the labour force and/or gaining most of their income from benefits, into employment. The peak at $180 a week corresponds approximately to the gross value of the unemployment and sickness benefits for single adults.[8] The increase in density in the mid-to-high income ranges could plausibly be due to transitions into full-time employment, real income increases for those in full-time employment, or a combination of both.

One useful insight from the analysis is that the process of economic change does not seem to have led to an increased concentration of Māori on low weekly incomes—if we define 'low' in absolute terms, for example as less than $500 a week (which is equivalent to $26,000 a year). Instead, the growth was fairly well spread across income levels above $500.

Changes in the income distributions of Māori males and females are plotted in Figures 3 and 4 respectively. The male distribution shows an increase in the concentration of people in the range of $650–$800 a week (around the central peak). This change in shape suggests that a decline in the dispersion of male incomes is likely to have occurred.[9] The change in the female distribution, on the other hand, could be characterised more simply as essentially a rightward shift, with less sign of change in shape or dispersion.

[8] While many respondents report their benefit income in net terms, SNZ converts these values to the gross equivalent.
[9] Measures of dispersion, such as the 90/10 percentile ratio, confirm that this was the case.

Figures 5–7 depict the changes in the income distribution for employed Māori, employed Māori males, and employed Māori females. These changes appear fairly undramatic. There is a noticeable reduction in the density of the income distribution function at income levels below the modal point and a thickening of densities at various income levels above the modal point, as one would expect given the recorded growth in real incomes.

The number of Māori respondents in the Income Survey sample is relatively small. The estimates given here for employed Māori, for example, draw on samples of around 2,600 in 1997/98 and 3,400 in 2002/03. The estimates for employed men and women are based on samples of roughly three-fifths that size. As discussed in Section 3, the Income Survey also has a number of design features that reduce the accuracy of its estimates, compared with estimates from a simple random sample. It is worth asking, therefore, what confidence can be attached to these kernel density estimates of the income distribution changes.

Figure 8 repeats the initial estimates of the working-aged Māori income distribution in 1997/98 and 2002/03, this time with 95% confidence intervals plotted around the 1997/98 income distribution function. Although we do not show them, confidence intervals on the 2002/03 distribution could also be plotted. A statistically significant change would be denoted by a gap between the two sets of bands. In fact, the confidence intervals shown in Figure 8 encompass both the 1997/98 and the 2002/03 estimates at most points of the distribution, indicating that the changes recorded during this five-year period are generally not statistically significant. Three points of change are clearly outside the confidence intervals for 1997/98: the decline in the proportion of the sample with zero incomes; the decline in the 'mini-peak' of people with incomes of around \$180 per week; and the rise in the central peak of the distribution, corresponding to incomes of around \$550 a week.

To pin down the significant changes more accurately, Figure 9 plots the *difference* between the kernel density estimates for the base and end periods, and 95% confidence intervals around that difference. Significant change occurs when both confidence interval bands are above, or below, zero. Figure 9 tells a similar story to Figure 8, identifying the same regions of growth and decline: a decline in the proportion of the sample with zero incomes; a decline in the ‘mini-peak’ of people with incomes of around $180 per week; and a rise in the central peak of the distribution, covering incomes of between $500 and $600. There are also pockets of significant growth in the density of the income distribution at higher levels of income, such as around $700–870 per week, $1,100–$1,500 per week and $1,900–2,500 per week.[10]

Figures 10 and 11 give analogous sampling error information for the income distribution of *employed* Māori. Although the direction of change is clearly one of redistribution towards income levels of $500 and above, significant differences are harder to identify. Figure 11 suggests significant growth in the proportion of employed Māori with incomes of $1,250–$1,450 and $1,900–$2,200 per week. However, the distance of confidence intervals from zero is small, suggesting these changes could fail a more precise test of significance.

Summarising these results, there is evidence that the weekly income distribution for working-aged Māori became more regular in shape between 1997/98 and 2002/03. There were significant reductions in the proportion of people with zero incomes and incomes in the range of $150–200 a week (measured in June 2000 dollars). There were some significant increases in the proportion with incomes at or above the central peak. Those increases were most concentrated in the following income bands: between $500 and $600, between $1,100 and $1,500, and between $1,900 and $2,500.

Sampling errors are larger if one focuses solely on employed Māori, and the income changes for this sub-population were on average smaller. As a result, the precise changes in the income distribution for employed Māori are more difficult to identify. However, the data indicate statistically significant increases in the proportion of employed Māori with incomes in the ranges of $1,250–$1,450 and $1,900–$2,200 per week.

[10] Note that we are using a fairly conservative approach to estimating sampling errors and confidence intervals, as discussed in Appendix B.

5.2 Summary measures of income change and income inequality

More conventional measures of the changes that occurred at specific points of the Māori income distribution, and the impact of those changes on dispersion, are presented in this section. The upper part of Table 6 gives results for all working-aged Māori (both genders combined), while the lower section gives results for employed Māori. Table 7 gives equivalent estimates for the European/Pākehā population. Estimates of changes that are significant at 95% confidence level are marked with an asterisk.

There were substantial increases in the dollar-value level of most percentiles of the working-aged Māori income distribution, from the 10th to the 90th. While a 10–12% increase was typical, the 25th percentile gained 22%. These large increases probably reflect at least in part the impact of the redistribution of more than 2% of Māori from zero to positive incomes, causing an upward shift in the position of all percentile rankings above zero (those falling in the 'positive incomes' part of the distribution). The size of the movement in each percentile of the distribution would also have been influenced by the shape of the income distribution at that point.[11]

Three different measures of dispersion are shown in the upper section of Table 6: the Gini coefficient of income inequality; the standard deviation of log incomes; and the inter-quartile range in the log income distribution. All measures of income inequality declined in this period (although only the decline in the inter-quartile range was statistically significant). This reduction in inequality is not surprising, given the growth in Māori employment rates and big reductions in the fraction of Māori with zero incomes or incomes from benefits.

The lower section of Table 6 gives results on income changes for employed Māori, a group whose incomes were less directly affected by the large increases in employment rates that occurred in this period. The income changes

[11] For example, percentiles at the lower and upper tails of the income distribution are more widely spaced simply because fewer people are located in these regions (the density of the income distribution function is lower), and therefore a rightward shift of all percentile rankings is likely to have a greater impact on the income levels corresponding to these 'extreme' percentiles.

experienced by employed Māori were more modest. All parts of the income distribution experienced some real income growth, ranging from 4.4% at the median to 7% at the 90th percentile and 11% at the 10th percentile. The increases were somewhat larger at the upper and lower ends than at the centre of the distribution. The changes in the mean and median are significant at the 95% confidence level. The changes in other percentiles are (marginally) insignificant at this level of significance.

The net effect of this pattern of change was relatively little change in overall income inequality among employed Māori. The Gini coefficient increased slightly, while the standard deviation of log income and the inter-quartile range declined slightly. The 90/50 and 50/10 percentile ratios suggest that the lower half of the income distribution for employed Māori became somewhat more compressed, while the upper half became somewhat more spread out (as suggested by the pattern of increases by percentile). However, none of the changes in the dispersion indices were statistically significant.

It is worth noting that these results do not give any direct indication of the income growth that would have been experienced by individuals who were employed throughout the period of study. The population of employed Māori expanded in this period and its composition is likely to have changed as a result, affecting the amount of income growth at each percentile in the distribution. Only a longitudinal dataset could accurately measure the income growth experienced by individuals.

5.3 Factors contributing to the income distribution changes

It is useful to think of changes in the income distribution for individuals in a given population (in this case, working-aged Māori) as the product of several sets of factors:

- Changes in the level or distribution of individual attributes that are related to income, such as age, residential patterns, family size, and educational attainment. These individual attributes may be correlated with income because they are differently rewarded in the labour market;[12] because they are correlated with differences in labour supply behaviour; or because they are linked to differences in entitlements to government income support.

- Changes in levels and patterns of labour market activity, such as the employment rate or the number of hours worked.

- Changes in the wage structure (representing the level and distribution of rewards that can be earned in the labour market for different skills and attributes).

- Changes in the level or distribution of unearned income, such as income support payments from the government.

Previous research on average Māori incomes has shown that age, education, rural/urban locality, hours of work, and occupation are all significant predictors of individual income level in an income regression context (Maani, 2000; 2002). These attributes and dimensions of labour market activity are therefore natural candidates for explaining changes in the Māori income distribution. As noted in Section 4.1, the period of study was marked by large increases in the total employment and full-time employment rates of Māori; a small upward shift in the age structure; increases in the proportions of Māori with higher levels of school and/or post-school qualifications; and a shift in the employment structure towards non-manual and more highly skilled occupations. There was little change in average hours worked.

In this section, an attempt is made to identify the role played by changes in the characteristics of Māori between 1997/98 and 2002/03 in bringing about income distribution changes, using techniques developed by DiNardo, Fortin and Lemieux (1996).[13] Briefly described, the following approach is taken. To estimate the effect of the change in a particular population characteristic, such as the age structure, the base period (1997/98) sample of working-aged Māori is reweighted

---

[12] Or correlated with other unmeasured attributes that are differentially rewarded in the labour market.

[13] For other recent applications of the DiNardo, Fortin and Lemieux method, see Hyslop and Maré, 2001; Wilkins, 2003; and Barsky et al, 2001.

so that the distribution of age in that sample matches the distribution that existed by the end of the study period (in 2002/03). The income distribution associated with this age-reweighted 'counterfactual' is then obtained and compared with the actual start and end period income distributions. The goal is to estimate the amount and pattern of change in the density of incomes that could potentially be explained by the change in age structure.

More specifically, in the DiNardo, Fortin and Lemieux approach, each observation is viewed as a vector (y_i, x_i, t_i) consisting of an income y, a vector of characteristics x and a date t, and belongs to a joint distribution $F(y,x,t)$ of incomes, characteristics and dates. The density of incomes at a particular point in time $f_t(y)$ can be written as the integral of the density of income conditional upon a set of individual characteristics and a date t_j, over the distribution of individual characteristics $F(x|t_x)$ at date t_x:

$$f_t(y) = \int f(y \mid x, t_j = t) dF(x \mid t_x = t)$$

$$\equiv f(y; t_j = t, t_x = t)$$

This notation allows us to express equations for counterfactual densities, with t_i denoting the date from which the function mapping characteristics to incomes is drawn, and t_x denoting the date from which the distribution of characteristics is drawn. For example, while $f(y;\ t_j=98,\ t_x=98)$ represents the actual density of incomes in 1997/98, $f(y;\ t_j=98,\ t_x=03)$ represents the density that would have resulted in 1997/98 if characteristics were as observed in 2002/03. This hypothetical density can be identified as follows:

$$f(y;\ t_y=98,\ t_x=03) = \int f(y|x,t_y=98)dF(x|t_x=03)$$

$$= \int f(y|x,t_y=98)\,\psi_x(x)dF(x|t_x=98)$$

where $\psi(_x)$ is a reweighting function:

$$\psi_x(x)= dF(x|t_x=03)/\ dF(x|t_x=98).$$

The counterfactual income density is identical to the 1997/98 density except for the reweighting function, so once an estimate of $\psi_x(x)$ is obtained, the counterfactual density can be estimated by weighted kernel methods such as:

$$f(y;\, t_j{=}98,\, t_x{=}03) = \sum_{i=1}^{n} \frac{\theta_i}{h} \hat{\psi}_x(x) K\left[\frac{w - w_i}{h}\right]$$

where the summation is over observations in the 1997/98 sample. This is simply a weighted version of the kernel density equation on page 20. Similarly, any summary measure such as the median or Gini coefficient can be calculated as a weighted median or weighted Gini coefficient. Essentially, each individual in the 1997/98 is reweighted so as to give the same distribution of characteristics as in the 2002/03 sample.

Applying Bayes' rule to the ratio $dF(x|t_x{=}03)/\ dF(x|t_x{=}98)$ gives the following reweighting function:

$$\psi_{xi}(x) = .\frac{\Pr_i(t_x = 03 \mid x)}{\Pr_i(t_x = 98 \mid x)} . \frac{\Pr_i(t_x = 98)}{\Pr_i(t_x = 03)}$$

where $\Pr_i(t_x{=}03|x)$ is the conditional probability that an individual with attributes x is observed in 2002/03 and $\Pr_i(t_x{=}03)$ is the unconditional probability of being observed in that year. To obtain estimates of the reweighting function, for each individual i, we pool the individuals in the two sample periods and estimate the probability that individual i is observed in 2002/03, given attributes x, using a logit model for the binary dependent variable t. We then use the estimates from this model to predict, for each individual observed in 1997/98, the relative probability that he/she would be observed in 2002/03 versus 1997/98 (the first term of the equation above) and adjust this by the sample proportions, $(\Pr_i(t{=}98)/\Pr_i(t{=}03)$. Intuitively, this reweighting scheme puts more weight on households with attributes that are more likely to occur in 2002/03 and less likely to occur in 1997/98.

Age and education-based counterfactual income distributions are shown in Figure 12. The line plotted for age, for example, represents the change in the density of the income distribution that could be expected to occur on the basis of the ageing of the population between 1997/98 and 2002/03. Similarly, the line plotted for education represents the estimated effect of the rise in educational attainment on the income density. Both of these effects are calculated under the assumption that the relative incomes for different values of the attribute (age and education respectively) remain as they were in 1997/98. (Stated another way, the wage structure and benefit structure are held constant.) The lines for age and education can be compared with the bold line representing the *actual* total change in income distribution that was recorded between the two points in time.[14]

The results in Figure 12 suggest that population ageing is likely to have had a very minor impact on the distribution of income among working-aged Māori. The impact of educational change was potentially more substantial, however. The estimated counterfactual income distribution taking educational change into account shows a reduction in the proportion of people at low income levels and an increase in the proportion above a threshold of about $450 per week.

Counterfactuals illustrating the effects of increased labour market activity are shown in Figure 13. The employment rate effect counterfactual simply captures the change in the proportions of Māori who were employed full-time and employed part-time. This effect is large, and appears capable of 'explaining' at least two-thirds of the total actual change in the Māori income distribution between 1997/98 and 2002/03. This employment rate counterfactual 'predicts' much of the decline in the proportion of Māori within incomes in the $120–220 range, as well as much of the increase in the proportion within incomes above $500 a week.

A counterfactual income distribution that incorporates information on the change in the distribution of hours worked *and* the distribution of employed Māori across occupations, as well as labour force status, is also plotted in Figure 13. Occupations are defined at 2-digit level. The combined impact of these three dimensions of labour market activity is slightly greater than the effect of labour force status alone, and 'explains' a slightly greater proportion of the total change in the income density function.

[14] Note that each factor is considered separately: the counterfactuals are not cumulative.

Finally, a multivariate or 'full model' counterfactual is also plotted in Figure 13. This incorporates all of the factors considered so far: age structure, education, employment rates, hours, and occupational structure. Note that these factors are highly correlated with each other, and therefore the combined effect is much less than the sum of the individual contributions. The 'full' attribute model accounts for somewhat more of the total change in income densities than the labour market activity counterfactual, particularly at higher levels of income.

Table 8 gives numerical information on the counterfactual income distributions, showing estimated income levels at the 10th, 25th, 50th, 75th, and 90th percentiles of income; the relative contribution of each counterfactual to the total income change that was recorded at each percentile; and whether the estimated income change component is statistically significant at that point in the income distribution. The effects of age distribution and educational level are basically confined to the upper half of the income distribution. Employment status, on the other hand, has its largest impact on the position of the 10th percentile, and a rather smaller impact on the position of the 90th percentile. The effect of the combined model, which includes both demographic and labour market variables, is more evenly distributed across the income spectrum. The combined model accounts for an estimated 62% of the actual total change in mean incomes; nearly all of the change in the 10th percentile, and about two-thirds of the change in the 90th percentile.

These simulations offer evidence in support of the view that the rise in the Māori employment rate was the single most important force transforming the income distribution, by shifting people from income levels below $500 per week to levels above this threshold. Changes in the distribution of other income-related attributes, particularly education, and changes in the occupational employment structure of Māori, may also have made a significant contribution. The combined effects of the changes in these measured attributes and labour market outcomes can account for at least two-thirds of the total change in the income distribution, and more than two-thirds at some specific points.

No attempt is made here to analyse the sources of change in the income distribution of *employed* Māori because of the imprecision with which the changes in that income distribution are measured.

5.4 Comparison of Māori and European/Pākehā income distribution changes

We turn now to a brief comparison of the changes in income distribution that were recorded for Māori with those recorded for European/Pākehā. The purpose is to identify the extent to which the changes affecting Māori were peculiar to the Māori ethnic group or more general in their impact. A secondary objective is to compare the quality of the estimates for Māori with those that can be obtained for Europeans, using the larger sample of Europeans in the dataset.

Tables 2–4 incorporate data for European/Pākehā in their right-hand columns. The employment rate for working-aged Europeans increased by 4 percentage points between 1997/98 and 2002/03, a smaller change in total level of activity than that recorded for Māori in this period. The proportion of Europeans who received no income in the survey reference week declined by 2 percentage points. The proportion receiving benefit income also declined. Shifts in the mean age, qualifications profile, and occupational profile of working-aged Pākehā were broadly similar to those recorded for Māori, although typically a little smaller in magnitude.

The average weekly income of all working-aged Pākehā/Europeans was 13% higher at the end of the period than at the beginning, and the average weekly income of employed Pākehā 8.7% higher. The increase in the average weekly income of all working-aged Pākehā was less than the comparable increase for working-aged Māori, leading to a decline in this particular ethnic income gap. However, there was little difference between the two ethnic groups in the average income growth rate for employed people, suggesting that patterns of job growth and wage increases in this period did not especially favour (or disadvantage) Māori.

Figures 14 and 15 compare the changes in the income distribution of all working-aged Māori with the changes in the European distribution. The latter is centred around a higher level of weekly income and has a higher peak at zero, but less of a 'blip' in the distribution at the $150–$200 level. The European distribution shows some similar processes of change to the Māori income distribution, including a decline in the proportion of individuals who are stacked at zero incomes; a rightward shift of the modal income level; a small reduction in the proportion earning incomes below the mode; and an increase in the proportion earning incomes in the range from $600 to $1,800 per week (measured in June 2000 dollars).

Income distributions for employed Māori and Pākeha are compared in Figure 16. The European/Pākehā distribution is once again centred around a higher level of weekly income and is significantly more dispersed (the central peak is lower). But the income distributions of both ethnic groups changed in a broadly similar manner, losing density at income levels below the central peak and gaining density at income levels above it.

Summary statistics on changes in the European/Pākehā income distribution are given in Table 7. The magnitude of the increases in income levels corresponding to the 10th, 25th, 50th, 75th, and 90th percentiles is reasonably similar to the rates of growth observed for Māori. The pattern of change in dispersion indices is also similar. The distribution of income across *all* working-aged Pākehā became significantly less unequal in this period. The inequality index measures for *employed* Pākehā suggest a small reduction in the dispersion of the bottom half of the income distribution, and a small increase in the dispersion of the upper half.

Standard errors on the European/Pākehā estimates are smaller, with the result that most of the changes recorded are estimated with significance. Interestingly however, the changes in the *dispersion* of income among employed Europeans were too small to be significant, like those for employed Māori. This also suggests that income distribution changes in this period were driven more by the transfer of people from unemployment or inactivity to employment, than by large changes in the structure of 'returns' to employment.

6 Summary and conclusion

This paper began by mapping out some of the main changes that were recorded in the income-related attributes and labour market activity patterns of working-aged Māori during the past five years—on the assumption that changes in these factors are important drivers of changes in the income distribution for any population or group. The Māori working-aged population grew older, with its average age increasing by 1.2 years, and experienced a significant increase in average levels of educational attainment. The occupational profile of working Māori shifted towards a greater share of skilled managerial, professional, and technical jobs, suggesting new patterns of job-related skill acquisition. Most significantly, there was a 9 percentage point increase in the employment rate of working-aged Māori (8 percentage points for men and almost 12 percentage points for women). All of these changes could potentially have contributed to income growth, and to changes in the distribution of incomes.

Income statistics for Māori show evidence of the types of changes that one would expect in a period of rapidly increasing employment. The proportion of working-aged Māori with zero incomes in the survey reference week declined from 10% to 8%, and the proportion reporting income from Work and Income benefits declined from 35% to 28%. The average real income gain between 1997/98 and 2002/03, averaged over all working-aged Māori, was 16%. The average real income gain for employed Māori was 8%.

Kernel density estimates of the density of the income distribution at different levels of income reveal a significant decline in the density 'peak' at zero incomes; a significant decline in density at incomes of $150–$200 per week; and significant increases in density at several regions of the distribution above $500 a week. Areas of growth in the density of the income distribution were fairly evenly spread across lower-middle to high income levels, indicating that the recent changes in Māori skills and economic activity patterns have not led to a new concentration of Māori on low or lower-middle incomes.

Summary measures of income inequality for the total population of working-aged Māori show a reduction in inequality in this period. This decline was driven by a decline in the proportion of Māori with zero or benefit-level incomes, and by an increase in the clustering of individuals at the central peak of the distribution. Summary measures of income inequality among *employed* Māori show little change in this period.

The key changes in the Māori income distribution were broadly similar to the changes in the Pākehā/European income distribution, suggesting the possibility of similar drivers or processes of change.

One of the objectives of the research was to evaluate the capacity of the Income Survey data to support an analysis of the causes of change in the Māori income distribution. In practice, the sampling errors on kernel density estimates of the Māori income distribution were quite large relative to the size of the changes recorded between 1997/98 and 2002/03. This meant that the observed changes tended to be statistically significant only at certain points or within certain income ranges. In general, the changes in the income distribution of all working-aged Māori were better identified and more likely to be statistically significant than the (smaller) changes in the income distribution of employed Māori. The Income Survey is likely to become an increasingly useful data source for research on income distribution changes as more years of data are collected, because income changes measured over a longer time period are likely to be larger and therefore more likely to dominate survey measurement errors.

An analysis of the key drivers of change in the income distribution indicated that the transition of Māori who were not working at the start of the period into employment, had a powerful impact on the aggregate income distribution, and could 'explain' the majority of the total change observed. Changes in the distribution of other income-related attributes, particularly education, and changes in the occupational structure of Māori employment, may also have made a significant contribution.

References

Barsky, Robert, John Bound, Kerwin Charles and Joe Lupin, 2001. "Accounting for the black–white wealth gap: A non-parametric approach," Working Paper 8466*, NBER*..

Benton, Richard. 2003. "The Study Areas: A general profile," in *Well-Being and Disparity in Tāmaki-Makaurau,* Vol. 5: Background studies. Ed. Richard Benton. Report prepared for Te Puni Kokiri, James Henare Māori Research Centre, University of Auckland, Auckland.

Chatterjee, Srikanta and Nripesh Podder. 2003. "Economic Inequality in Colour: Some ethnic dimensions of income distribution in New Zealand 1984–98." Paper presented at the Annual Conference of the New Zealand Association of Economists, Auckland.

DiNardo, J.; N. M. Fortin and T. Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A semiparametric approach." *Econometrica*, 64:5, pp. 1001–44.

Hyslop, Dean and Dave Maré. 2001. "Understanding changes in the distribution of household incomes in New Zealand between 1983–86 and 1995–98." Working Paper 2001/21, The Treasury, Wellington.

Maani, Sholeh. 2000. "Secondary and tertiary educational attainment and income levels for Māori and non-Māori over time." Working Paper 2000/18, The Treasury, Wellington.

Maani, Sholeh. 2002. "Education and Māori relative income levels over time: The mediating effect of occupation, industry, hours of work and locality." Working Paper 2002/17, The Treasury, Wellington.

New Zealand Institute for Economic Research. 2003. "Māori Economic Development." Report to Te Puni Kokiri, NZIER, Wellington.

Wilkins, Roger. 2003. "Immigrant and native-born earnings distributions in Australia: 1982–1996." *Australian Journal of Labour Economics*, 6:1, pp. 83–115.

Figures

Figure 1: Published HLFS estimates of the aggregate Māori and Pākehā/European employment rates

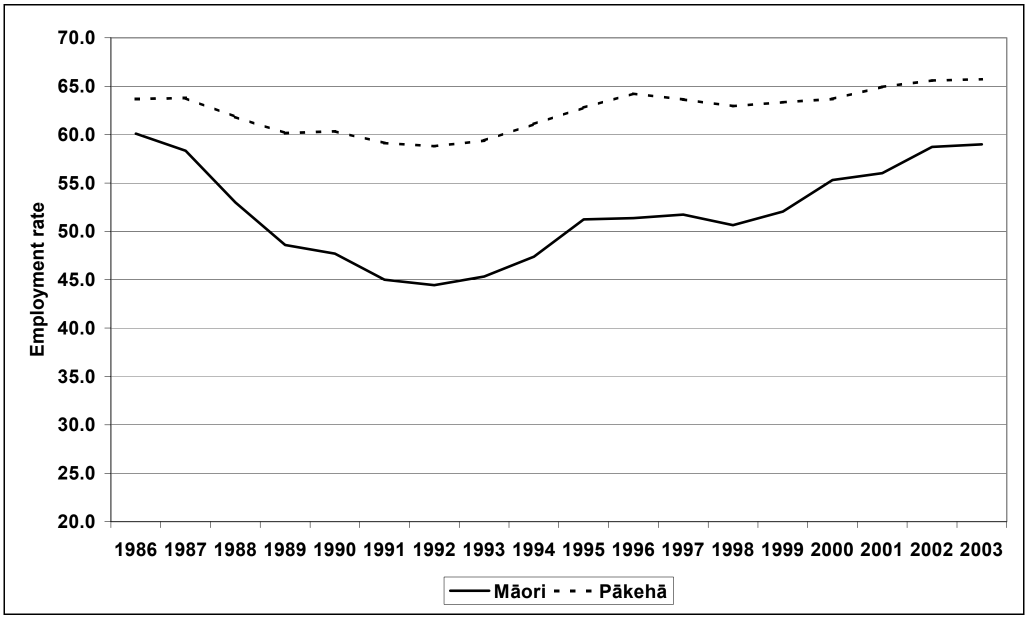


Figure 2: The log income distribution for all working-aged Māori in 1997/98 and 2002/03

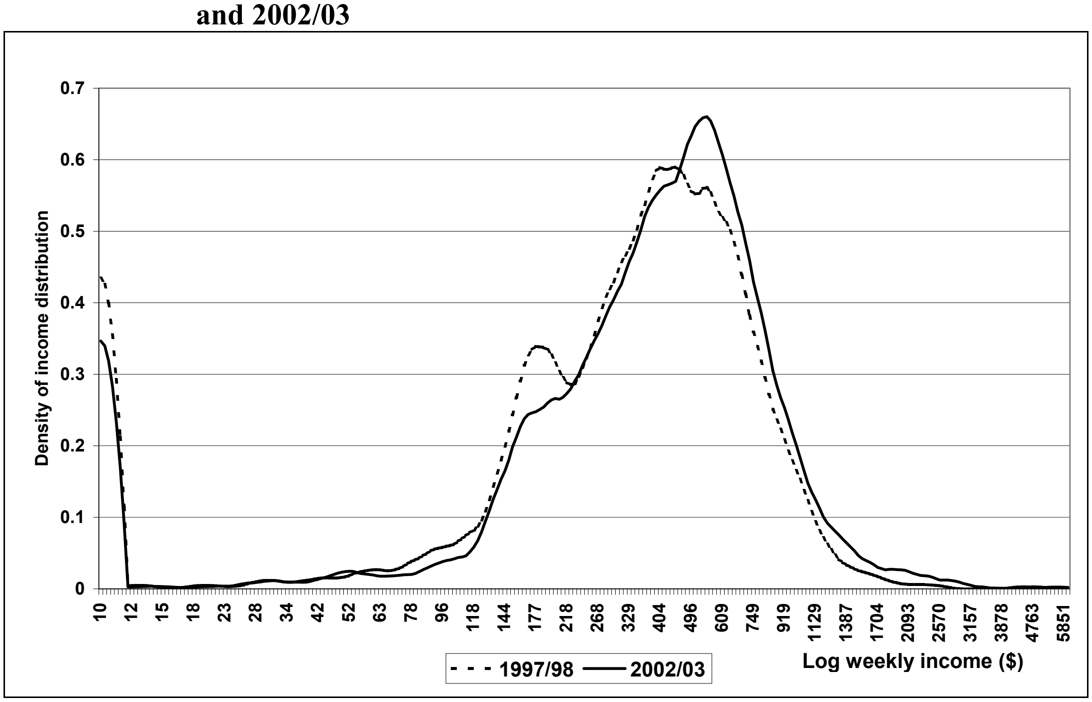


Figure 3: The log income distribution for working-aged Māori men in 1997/98 and 2002/03

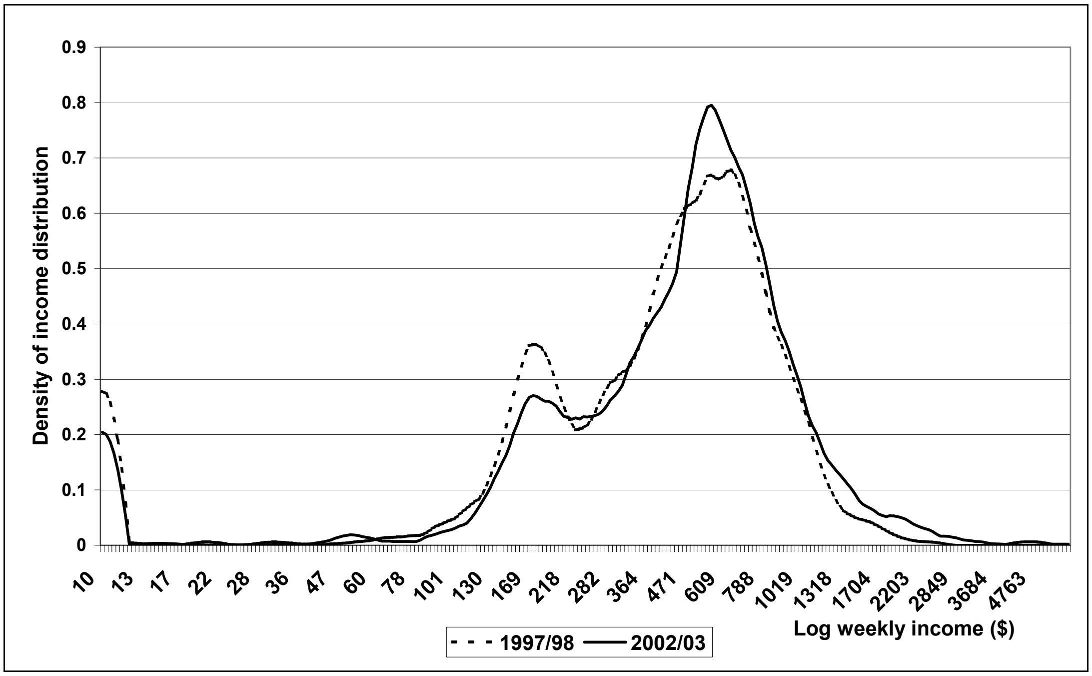


Figure 4: The log income distribution for working-aged Māori women in 1997/98 and 2002/03

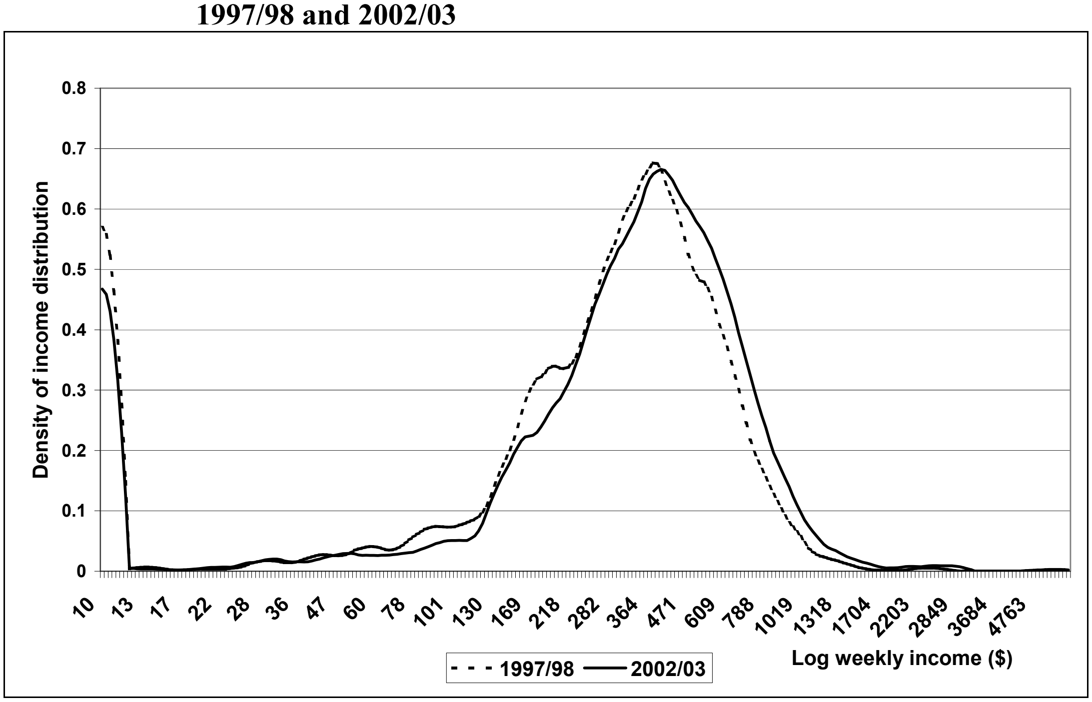


Figure 5: The log income distribution of employed Māori in 1997/98 and 2002/03

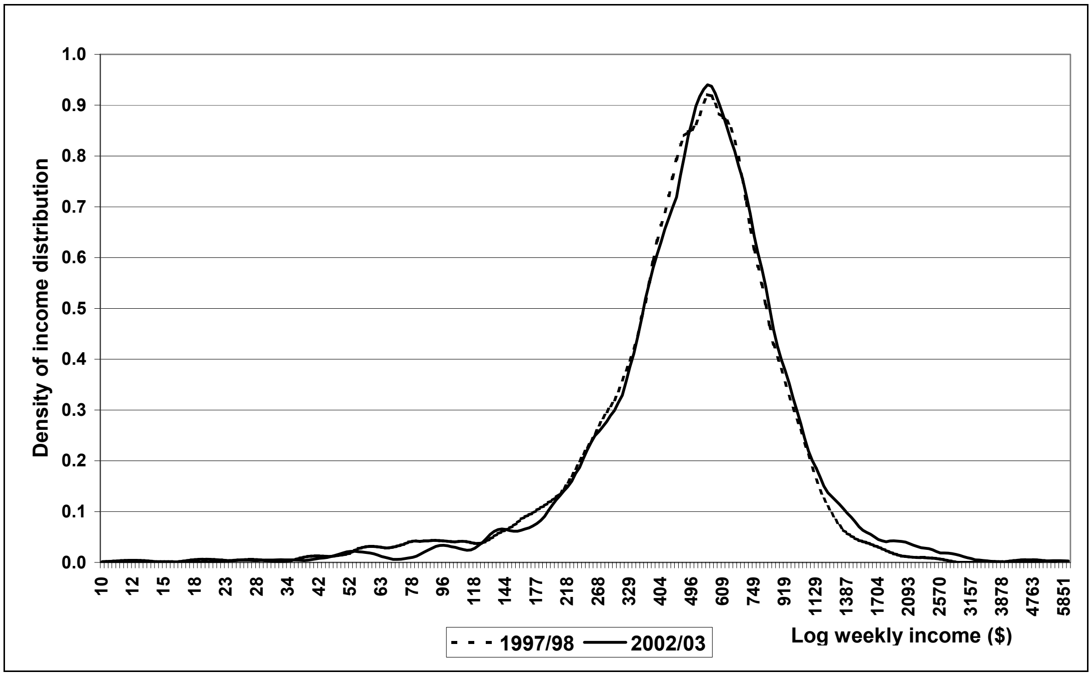


Figure 6: The log income distribution of employed Māori men in 1997/98 and 2002/03

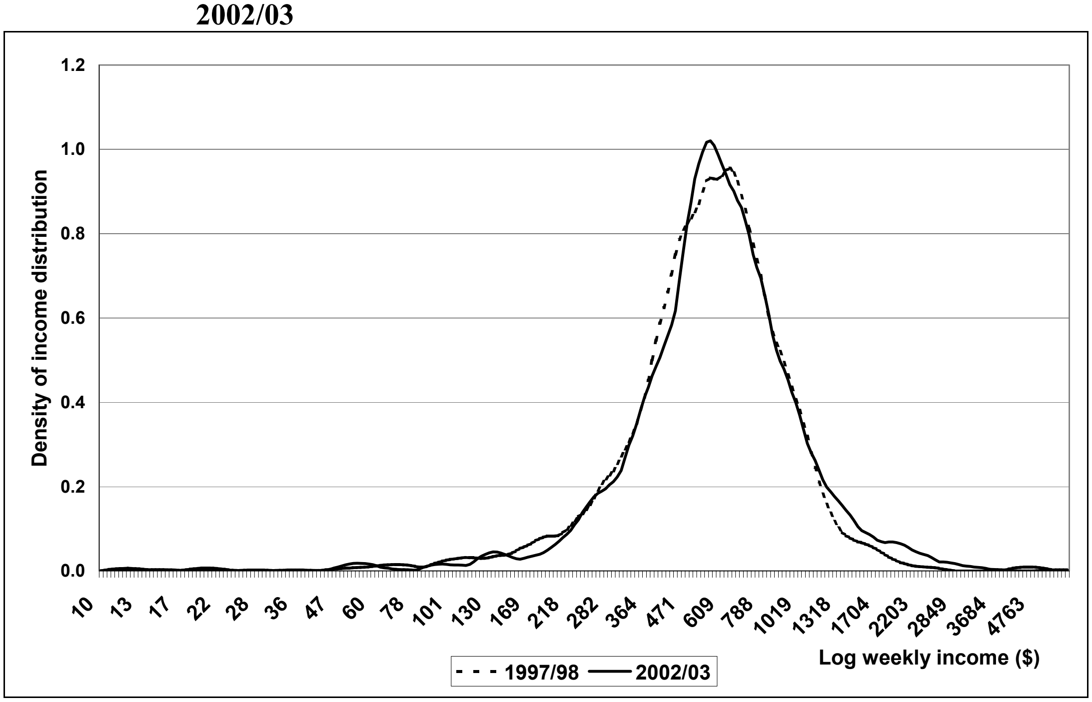


Figure 7: The log income distribution of employed Māori women in 1997/98 and 2002/03

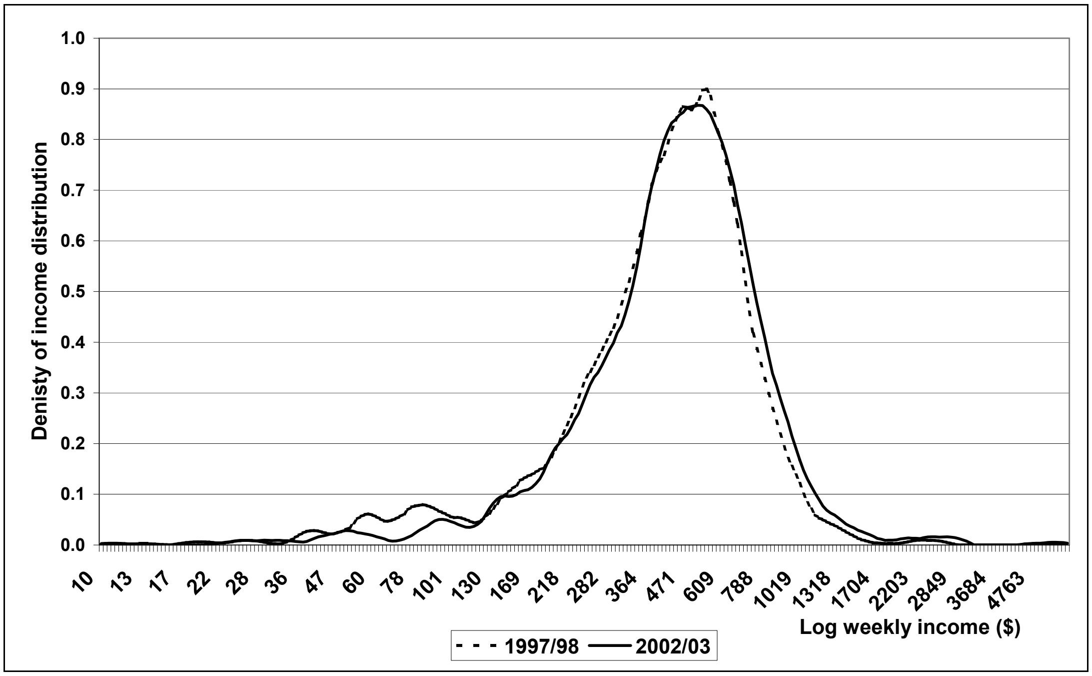


Figure 8: Confidence intervals on the log income distribution of working-aged Māori in 1997/98

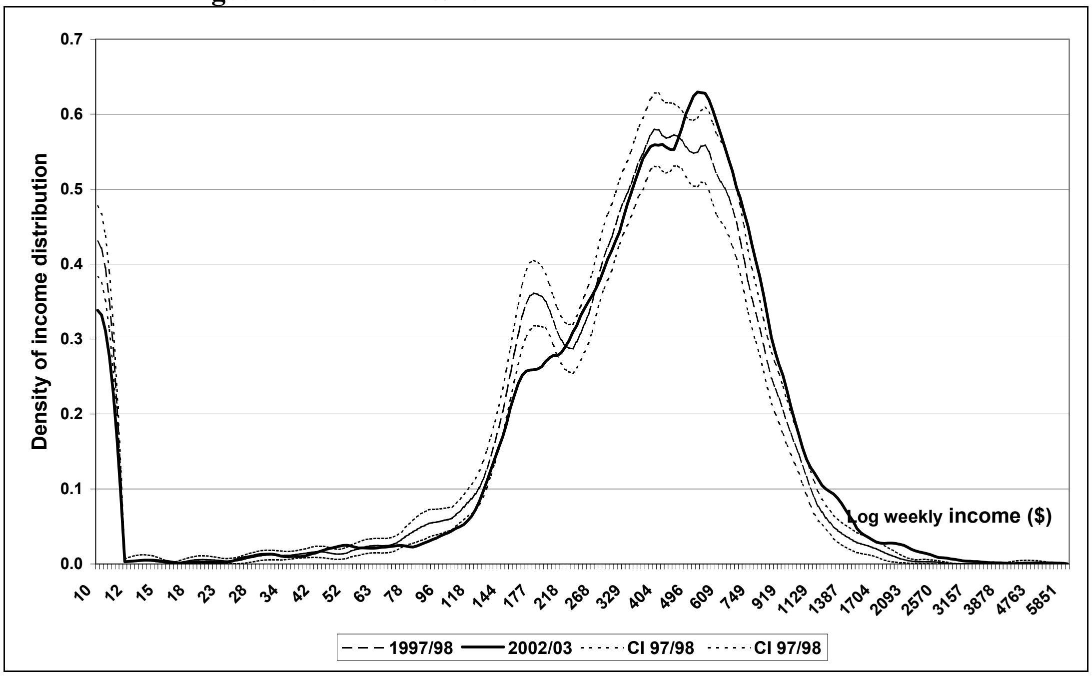


Figure 9: Confidence intervals on the change in the density of the log income distribution for working-aged Māori

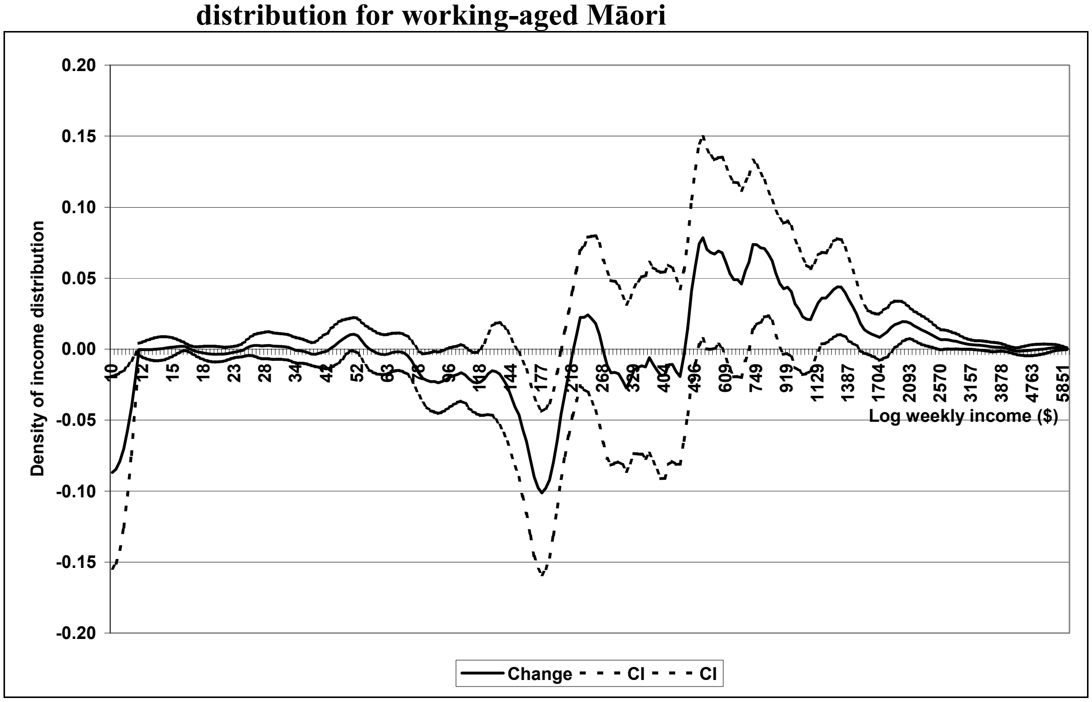


Figure 10: Confidence intervals on the log income distribution of employed Māori in 1997/98

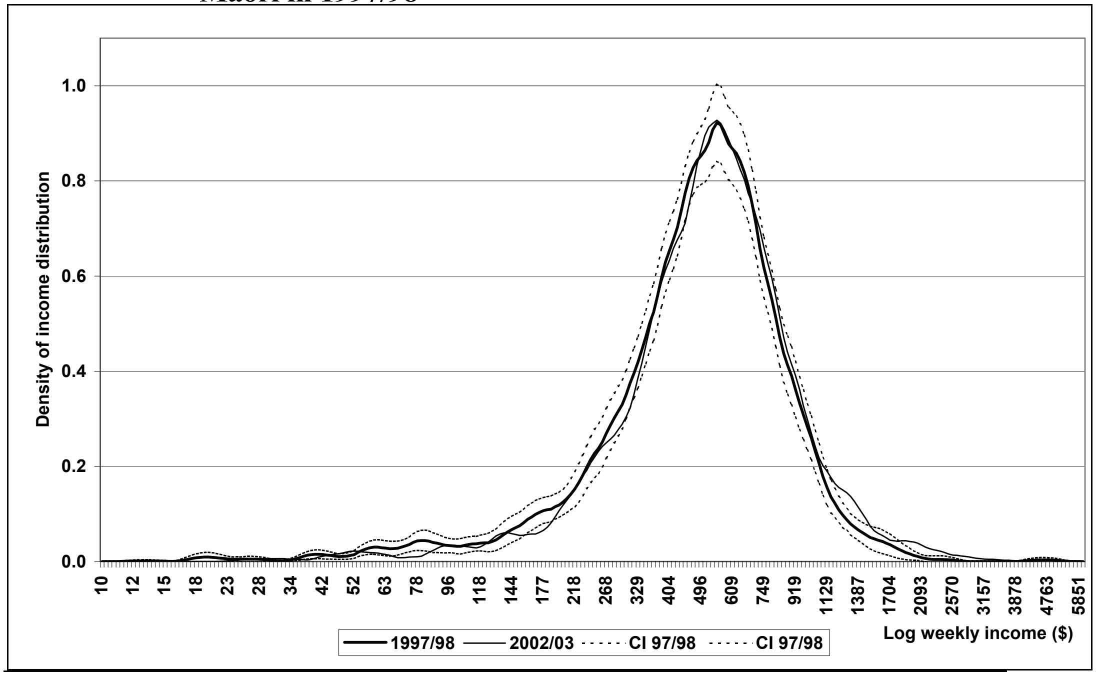


Figure 11: Confidence intervals on the change in the density of the log income distribution for employed Māori

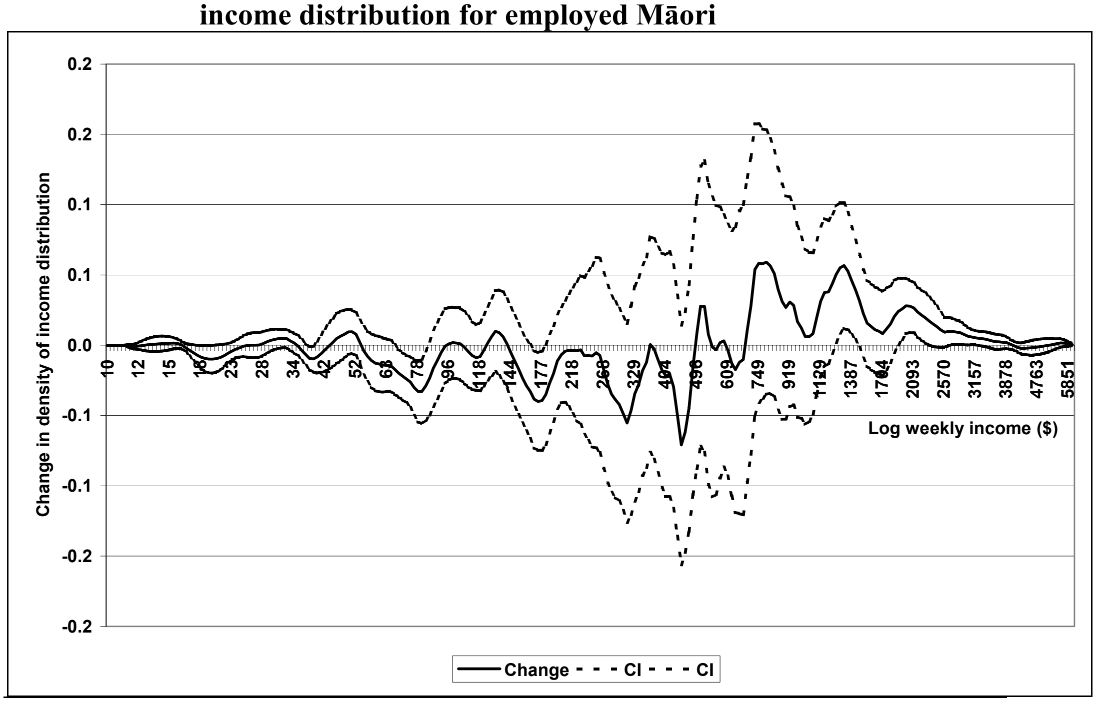


Figure 12: Age and education-based counterfactual income distributions

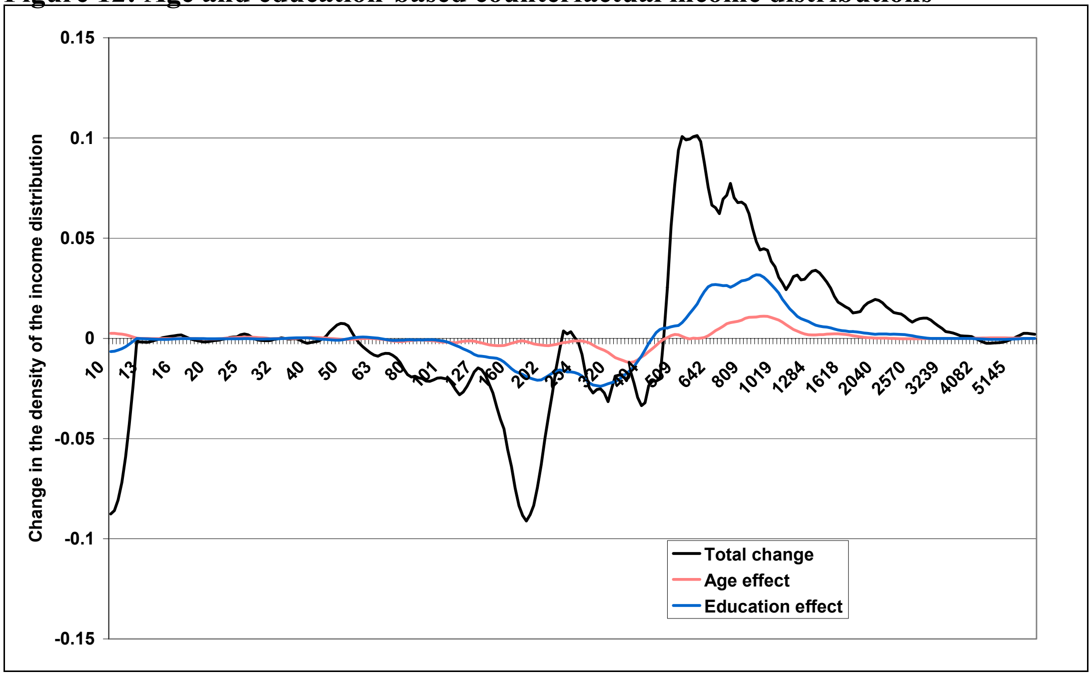


Figure 13: Employment rate and combined labour market activity counterfactual income distributions

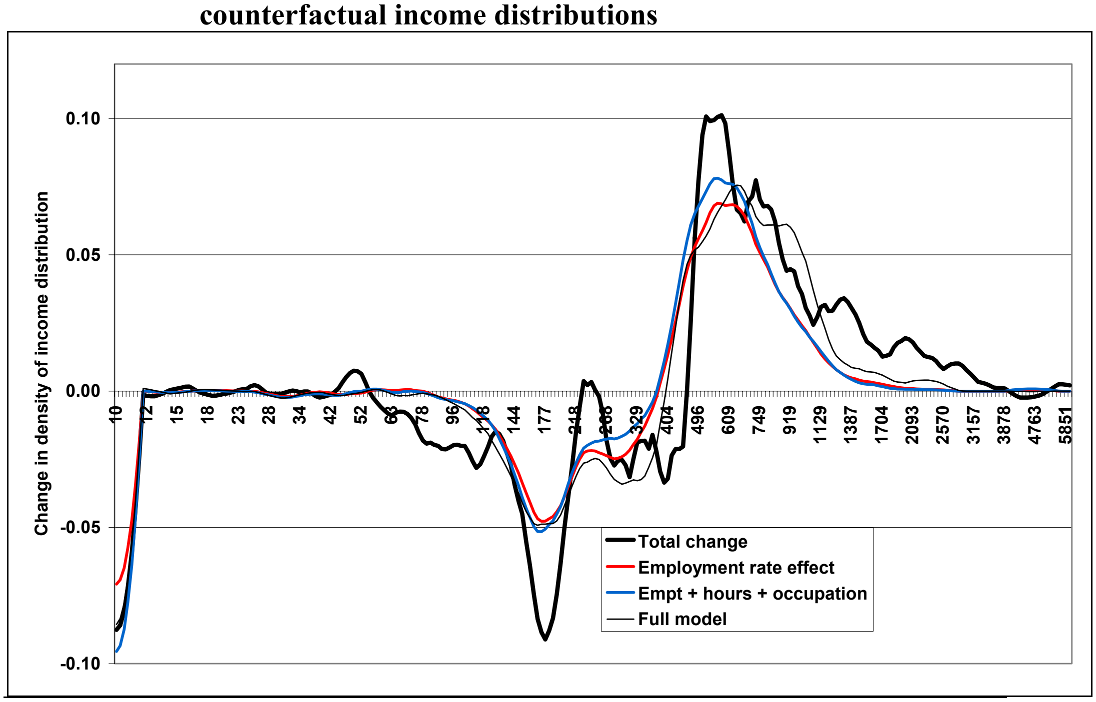


Figure 14: Māori and Pākehā log income distributions in 1997/98 and 2002/03

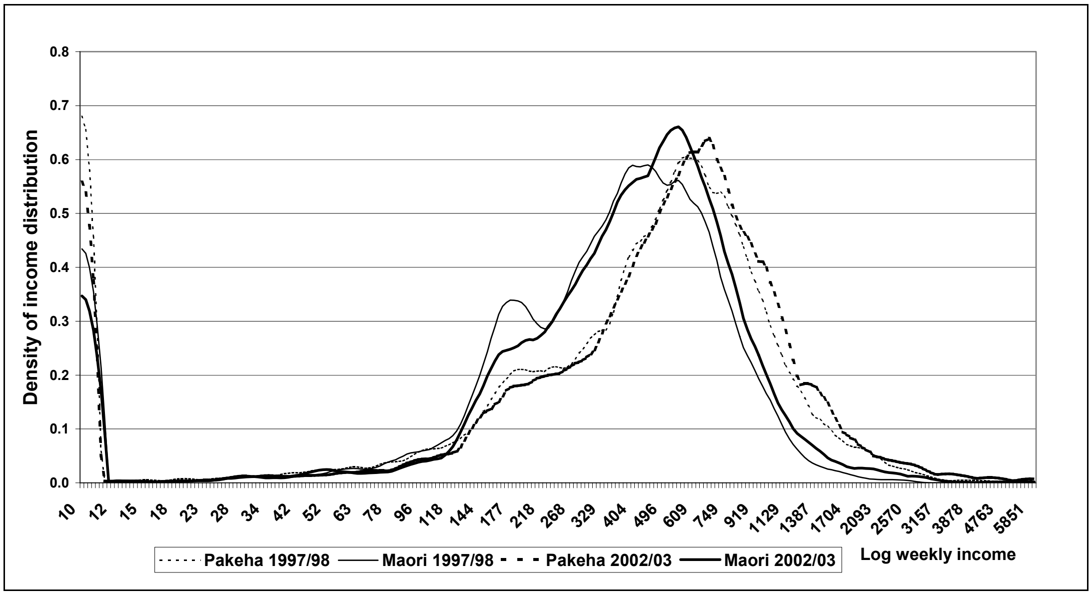


Figure 15: Comparison of the Māori and Pākeha income distribution changes

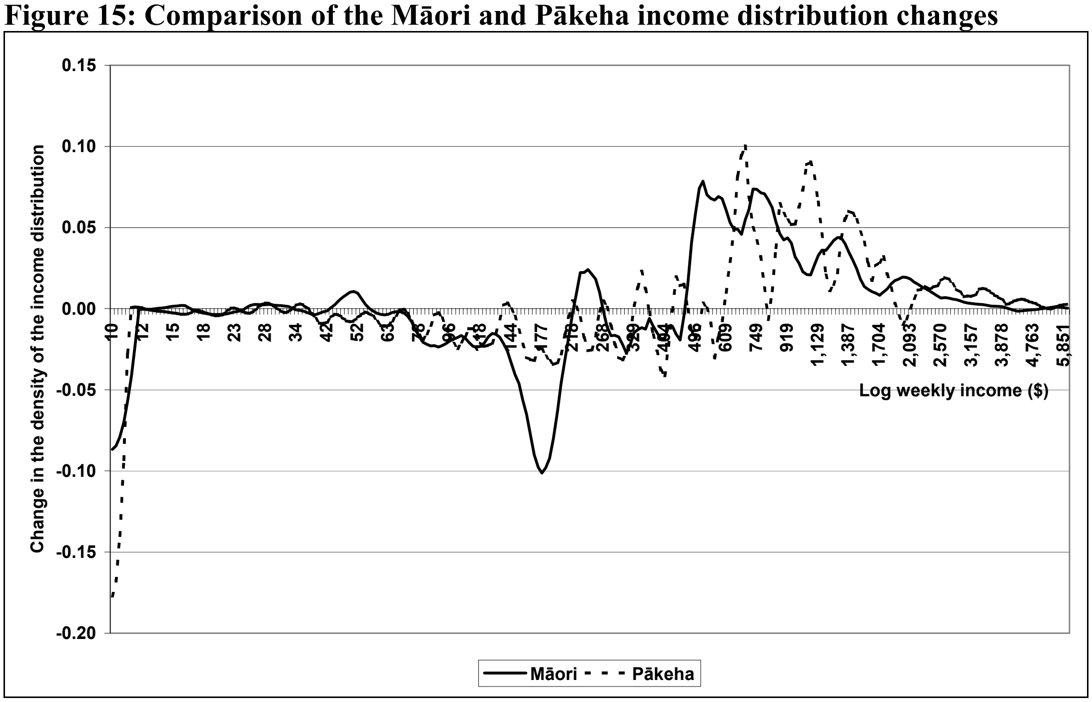
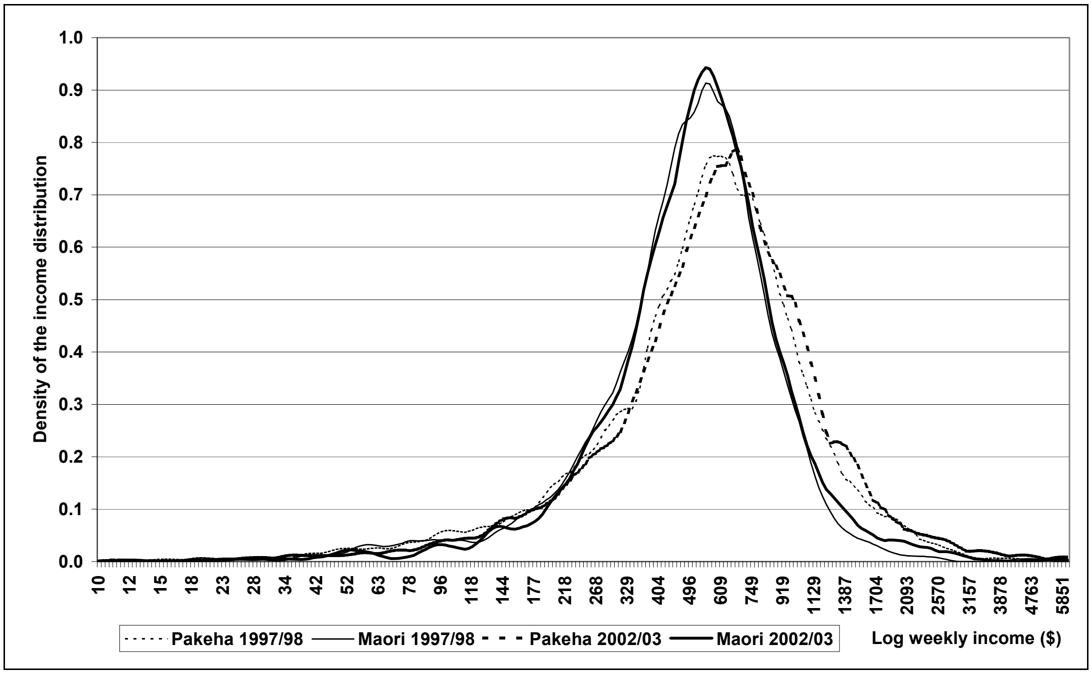


Figure 16: Log income distributions for employed Māori and Pākehā in 1997/98 and 2002/03



Tables

Table 1: Total sample sizes, including imputed records

	Māori			Pākehā/European		
	Males	Females	All	Males	Females	All
All aged 20-59						
1997	1,059	1,368	2,427	7,199	7,751	14,950
1998	1,027	1,356	2,383	7,112	7,731	14,843
2002	1,178	1,533	2,711	6,994	7,628	14,622
2003	1,102	1,491	2,593	6,898	7,476	14,374
Employed and aged 20-59						
1997	723	621	1,344	6,119	5,241	11,360
1998	667	611	1,278	5,934	5,091	11,025
2002	873	862	1,735	6,090	5,508	11,598
2003	836	822	1,658	6,054	5,428	11,482

Note: Imputed records were used in the base results but not in the estimation of errors.

Table 2: Imputation rates and sample sizes excluding imputed records

	Māori			Pākehā/European		
	Males	Females	All	Males	Females	All
Imputation rates						
All aged 20-59						
1997/98	22.9	16.2	19.1	20.0	13.0	16.4
2002/03	19.1	14.0	16.2	16.3	11.5	13.8
Employed sub-population						
1997/98	24.2	17.0	20.8	21.1	14.8	18.2
2002/03	21.4	16.2	18.8	17.2	13.1	15.2
Non-imputed sample sizes						
All aged 20-59						
1997/98	1,609	2,284	3,893	11,447	13,462	24,909
2002/03	1,845	2,601	4,446	11,626	13,363	24,989
Employed sub-population						
1997/98	1,054	1,022	2,076	9,511	8,802	18,313
2002/03	1,343	1,412	2,755	10,059	9,506	19,565

Table 3: Employment rate increases

	Māori			Pākehā/European		
	1997/98	2002/03	Change	1997/98	2002/03	Change
All working-aged						
Employment rate	57.4	66.8	*9.4*	76.5	80.7	*4.2*
Full-time employment rate	44.9	53.5	*8.6*	60.5	64.1	*3.7*
Males						
Employment rate	69.3	77.3	*8.1*	85.1	87.9	*2.7*
Full-time employment rate	61.4	69.8	*8.4*	78.3	81.1	*2.8*
Females						
Employment rate	47.2	57.9	*10.7*	67.9	73.5	*5.6*
Full-time employment rate	30.7	39.7	*9.0*	42.8	47.3	*4.5*

Table 4: Changes in income-related personal characteristics

	Māori			Pākehā/European		
	1997/98	2002/03	Change	1997/98	2002/03	Change
Mean age	35.0	36.2	*1.2*	38.7	39.7	*1.0*
Highest qualification (%)						
None	42.7	34.8	*-7.9*	20.4	16.8	*-3.6*
Lower school	11.1	11.5	*0.4*	11.0	10.5	*-0.5*
Higher school	9.6	9.2	*-0.4*	13.5	13.0	*-0.4*
Vocational	33.5	38.9	*5.4*	42.6	44.3	*1.7*
Degree	3.1	5.6	*2.5*	12.5	15.3	*2.8*
Occupations of the employed (%)						
Managerial	6.4	8.1	*1.7*	13.7	14.6	*0.9*
Professional	9.6	10.7	*1.1*	14.8	16.6	*1.8*
Technical	8.9	9.9	*0.9*	13.9	12.2	*-1.7*
Clerical	11.4	12.4	*1.1*	13.7	13.3	*-0.4*
Service and sales	14.5	14.4	*-0.1*	11.7	12.3	*0.5*
Agricultural	6.7	7.9	*1.2*	8.4	8.6	*0.2*
Trades	9.8	7.9	*-1.9*	10.3	10.5	*0.2*
Manual	18.6	17.2	*-1.3*	7.7	7.3	*-0.4*
Elementary	14.0	11.1	*-2.9*	5.6	4.4	*-1.2*
Not specified	0.2	0.3	*0.1*	0.2	0.1	*0.0*
Mean hours of the employed	38.0	37.8	*-0.2*	39.0	38.6	*-0.4*

Table 5: Summary of mean income changes

	Māori			Pākehā/European		
	1997/98	2002/03	Change	1997/98	2002/03	Change
Both genders						
Mean incomes, all aged 20-59 ($)	401	463	*15.6%*	535	604	*13.0%*
Mean incomes of the employed ($)	553	599	*8.2%*	664	721	*8.7%*
Mean incomes of the non-employed ($	195	191	*-2.2%*	115	115	*-0.2%*
Proportion with zero incomes (%)	10.2	8.1	*-2.1*	12.0	9.8	*-2.2*
Proportion with benefit income (%)	34.7	27.9	*-6.8*	11.9	10.3	*-1.6*
Males						
Mean incomes, all aged 20-59 ($)	488	557	*14.1%*	698	766	*9.7%*
Mean incomes of the employed ($)	623	670	*7.6%*	796	853	*7.2%*
Mean incomes of the non-employed ($	184	170	*-7.7%*	138	134	*-2.7%*
Proportion with zero incomes (%)	6.5	4.8	*-1.7*	6.0	5.0	*-1.0*
Proportion with benefit income (%)	24.3	19.2	*-5.1*	8.6	7	*-1.6*
Females						
Mean incomes, all aged 20-59 ($)	325	383	*18.1%*	372	443	*18.8%*
Mean incomes of the employed ($)	463	517	*11.5%*	499	564	*12.9%*
Mean incomes of the non-employed ($	200	200	*-0.2%*	104	106	*1.4%*
Proportion with zero incomes (%)	13.4	10.8	*-2.6*	17.9	14.6	*-3.3*
Proportion with benefit income (%)	43.8	35.4	*-8.4*	15.2	13.5	*-1.7*

Table 6: Summary measures of change in the Māori income distribution

	1997/98	2002/03	Change	% chge	1997/98 SE	2002/03 SE	SE on chge	Change signif at 95%
All working-aged Māori								
		A: Percentiles of the income distribution						
Mean	400.6	463.1	62.5	*15.6%*	6.0	6.2	8.6	*
90th	762.2	846.7	84.5	*11.1%*	18.1	16.9	25.5	*
75th	557.9	614.2	56.3	*10.1%*	11.7	9.1	14.1	*
50th	361.6	410.0	48.4	*13.4%*	5.7	7.1	9.2	*
25th	184.7	226.0	41.3	*22.3%*	3.2	5.4	6.6	*
10th	0.0	69.3	69.3			17.4	22.8	*
		B: Summary measures of income inequality						
Gini coefficient	0.402	0.395	-0.007	*-1.7%*	0.008	0.008	0.011	
Std dev log income	1.902	1.771	-0.131	*-6.9%*	0.045	0.044	0.067	
IQR (oflogs)	1.105	1.000	-0.105	*-9.5%*	0.025	0.022	0.035	*
Employed Māori								
		A: Percentiles of the income distribution						
Mean	552.9	598.5	45.6	*8.2%*	9.5	7.9	13.9	*
90th	894.0	954.9	60.9	*6.8%*	22.2	26.3	35.8	
75th	688.7	713.9	25.1	*3.7%*	11.2	7.6	14.4	
50th	514.5	537.3	22.9	*4.4%*	5.6	4.9	8.2	*
25th	367.5	385.0	17.5	*4.8%*	8.4	7.7	12.0	
10th	233.4	258.4	25.0	*10.7%*	8.6	10.6	13.2	
		B: Summary measures of income inequality						
Gini coefficient	0.280	0.293	0.013	*4.6%*	0.010	0.010	0.011	
Std dev log income	0.624	0.600	-0.024	*-3.9%*	0.019	0.024	0.032	
IQR (of logs)	0.631	0.617	-0.015	*-2.3%*	0.022	0.018	0.030	
90-50 (of logs)	0.546	0.579	0.033	*6.0%*	0.023	0.027	0.035	
50-10 (of logs)	0.792	0.723	-0.069	*-8.7%*	0.032	0.034	0.047	

Table 7: Summary measures of change in the European/Pākehā income distribution

	1997/98	2002/03	Change	% chge	1997/98 SE	2002/03 SE	SE on chge	Signif at 95%
All working-aged European/Pākehā								
		A: Percentiles of the income distribution						
Mean	534.6	604.0	69.4	*13.0%*	5.4	6.1	7.8	*
90th	1033.1	1131.0	98.0	*9.5%*	12.9	16.2	20.5	*
75th	723.1	791.7	68.6	*9.5%*	6.1	6.8	8.9	*
50th	467.7	520.4	52.7	*11.3%*	3.9	5.2	6.8	*
25th	202.2	253.5	51.3	*25.4%*	2.9	4.7	5.5	*
10th	0.0	11.6	11.6		0.0	10.8	10.5	
		B: Summary measures of income inequality						
Gini coefficient	0.449	0.438	-0.011	*-2.4%*	0.004	0.004	0.006	
Std dev log income	2.134	2.006	-0.128	*-6.0%*	0.019	0.020	0.029	*
IQR (of logs)	1.274	1.139	-0.135	*-10.6%*	0.016	0.017	0.024	*
Employed European/Pākehā								
		A: Percentiles of the income distribution						
Mean	663.7	721.2	57.4	*8.7%*	6.3	6.0	8.6	*
90th	1157.5	1252.0	94.5	*8.2%*	16.4	12.5	20.3	*
75th	813.1	882.4	69.3	*8.5%*	6.9	5.4	8.8	*
50th	579.2	615.4	36.2	*6.3%*	2.7	3.4	4.4	*
25th	387.1	408.6	21.5	*5.5%*	2.5	3.5	4.4	*
10th	210.9	237.4	26.5	*12.6%*	5.3	4.5	7.0	*
		B: Summary measures of income inequality						
Gini coefficient	0.345	0.351	0.006	*1.7%*	0.005	0.004	0.005	
Std dev log income	0.741	0.721	-0.020	*-2.7%*	0.009	0.007	0.012	
IQR (of logs)	0.742	0.770	0.028	*3.8%*	0.008	0.009	0.012	
90-50 (of logs)	0.693	0.708	0.015	*2.1%*	0.013	0.012	0.017	
50-10 (of logs)	1.009	0.953	-0.056	*-5.5%*	0.021	0.018	0.029	

Table 8: Counterfactual income distributions

	Actual	Counterfactual distributions					Actual
	1997/98	Age	Educn	Empt R	LM activity	Full model	2002/03
Percentiles of the income distribution ($)							
mean	400.6	404.6	415.6	429.5	435.1	441.2	466.0
P10	0.0	0.0	0.0	58.5	79.3	67.1	69.3
P25	184.7	184.9	190.5	206.6	216.3	215.1	226.0
P50	361.6	363.5	377.4	396.3	401.4	406.5	410.0
P75	557.9	568.2	584.7	594.4	596.0	608.5	614.2
P90	762.2	774.8	792.5	792.5	794.4	818.6	846.7
Change from base year=1997/98 ($)							
		Age	Educn	Empt R	LM activity	Full model	Actual
mean		4.1	15.1	28.9	34.5	40.6	65.5
P10		0.0	0.0	58.5	79.3	67.1	69.3
P25		0.2	5.9	21.9	31.6	30.4	41.3
P50		1.9	15.8	34.7	39.8	44.9	48.4
P75		10.3	26.9	36.5	38.2	50.6	56.3
P90		12.6	30.3	30.3	32.2	56.4	84.5
Estimated contribution to the total income change (% share)							
		Age	Educn	Empt R	LM activity	Full model	Actual
mean		6.2	23.0	44.2	52.8	62.0	100.0
P10		0.0	0.0	84.4	114.4	96.8	100.0
P25		0.6	14.2	53.1	76.7	73.6	100.0
P50		3.9	32.7	71.7	82.3	92.8	100.0
P75		18.3	47.7	64.9	67.8	89.8	100.0
P90		14.9	35.9	35.9	38.2	66.8	100.0
Estimated change significant at 5 percent error level							
		Age	Educn	Empt R	LM activity	Full model	Actual
mean		*	*	*	*	*	*
P10				*	*	*	*
P25				*	*	*	*
P50			*	*	*	*	*
P75		*	*	*	*	*	*
P90		*	*	*	*	*	*

Notes: Each counterfactual was estimated separately and the results shown reflect separate not cumulative effects.

Appendix A: Sampling error estimation

Survey design features

The HLFS-IS has several design features that affect the sampling errors of the statistics that are obtained from IS data. Those design features are functional in that they serve to maximise the accuracy of key survey estimates (in this case, key HLFS estimates such as the unemployment rate) for a given survey collection budget. However, if they are not taken into account in the calculation of standard errors, those standard errors are likely to be systematically smaller than they should be, and too much reliance could be placed on results that are due to sampling variability rather than true changes in the study population.

The HLFS, like many other surveys, has a clustered and stratified sample design (see Smith, 2001). It is clustered because selection into the sample occurs at the level of geographical units called primary sampling units (PSUs), each containing around 50–100 dwellings. Within PSUs, households are further divided into sub-groups called panels. Over time, all panels in a selected PSU, and all households within each panel, are incorporated into the survey sample. The selection process is also structured by an overall grouping of PSUs into 'strata', non-random sets of PSU that are grouped together according to various geographic and socio-economic variables. PSUs are selected independently within each strata.

Two further features of the HLFS that also affect the variance of the estimates are worthy of note. First, it is a rotating panel survey, in which households are interviewed for up to eight quarters before being dropped from the sample. Consequently, there is a (partial) overlap in the IS sample from one year to the next that should be taken into account: responses in 1998, for example, are not completely independent of responses in 1997 because of the sample overlap.

Second, the income responses of a substantial proportion of the sample—about 15% of all working-aged adults and 18% of working-aged Māori—are imputed by Statistics New Zealand because the individuals concerned could not be contacted to respond in person. Recall that the Income Survey is administered as a supplement to the HLFS each June quarter. The HLFS obtains responses from every adult in each household that is selected into the sample. Proxy responses to the core HLFS questions, given by other household members, are accepted if particular individuals cannot be contacted. However, SNZ does not take proxy responses to the Income Survey questions because they are considered to be too complex and personal to be accurately answered by others. Missing responses to the Income Survey are imputed instead, using the demographic and labour force information that *is* available for missing individuals and a hot deck imputation method.

Briefly described, the imputation procedure randomly selects an Income Survey response from the sub-set of respondents who matched the missing individual in their demographic and labour force characteristics, and assigns it to them. The variables used to form the donor classes for imputation are labour force status, self-employed versus employee, gender, age, highest qualification, full-time/part-time hours, ethnicity, and region of residence (Auckland versus other regions).

The imputation of records for IS non-respondents increases the accuracy of population estimates by increasing the representativeness of the sample.[15] However, because imputation does not bring in any new information, it does not increase the effective sample size for the purpose of calculating sampling errors. An adjustment must be made when standard errors are calculated. The usual method is to drop imputed cases from the sample and adjust the weights for the non-imputed cases to give population totals that are approximately correct. In this study, a reduced and reweighted dataset containing the non-imputed cases only was obtained from SNZ and used in the estimation of standard errors.

Replication methods for calculating sampling errors

There are two commonly-used replication methods for calculating variances and sampling errors: jackknife and bootstrap estimation. These replication-based methods are particularly useful for data from complex surveys, because the design features of such surveys make it more difficult to calculate

[15] Provided that people are missing for reasons that are not highly correlated with the imputation factors.

standard errors accurately using analytical methods. In this study, several versions of both jackknife and bootstrap estimation were tested and compared before a modified bootstrap method, described below, was adopted.

At its simplest, the jackknife method is implemented by successively dropping each observation (or sampling unit) from the sample and then re-estimating the statistic on the reduced sample. Each observation is dropped in turn, and then replaced. The variability of the statistic over the full set of replications provides a measure of the variance of the statistic.

Sample clustering can be taken into account though the manner in which observations are excluded from the sample. Shao and Tu (1995, p. 239) discuss a jackknife variance estimator in which each first-stage cluster (e.g. the PSU) is deleted from the sample in each replication (rather than the ultimate sampling unit, the household). SNZ uses a delete-a-group jackknife method to estimate HLFS and HLFS-IS sampling errors (see Smith, 2001). PSUs are formed into a list ordered by strata but randomised within strata. Groups (or sub-samples) are formed by systematically selecting PSUs down this list with a constant skip interval. Having divided the sample into *R* groups (e.g. 100), each of *R* replicate samples is formed by deleting one group from the sample. Weights are adjusted for each replicate sample to reflect the actual number of PSUs remaining in the sample, within each strata. The statistic of interest is recalculated *R* times, using each of the replicate samples and adjusted weights, and then a variance (or standard error) is calculated across estimates using a modified variance formula.

In the simplest version of the bootstrap method, *N* observations are selected, with replacement, from a dataset containing *N* observations. In that random drawing, some of the observations will appear once, some more than once, and some not at all. Using the new dataset, one recalculates the statistic of interest. This is done repeatedly, each time drawing a new random sample and re-estimating. The standard deviation of statistic across replications provides an estimate of the true standard error of the statistic.

This simple method of drawing samples for bootstrap replications can be modified to take into account survey design features such as clustering and stratification. Shao and Tu (1995, p. 246) propose an extension in which selection into each replication sample occurs at the level of the cluster, for example the PSU. Clusters are selected independently within each strata. Each bootstrap replicate sample is then formed by summing the sub-samples of PSUs across all strata. This ensures that the stratified structure of the sample is preserved.

Because full information on the identity of each strata and PSU within the HLFS-IS was not made available by SNZ for use in this study, a simplified version of the clustered and stratified bootstrap method described by Shao and Tu was implemented. We used the groupings (sub-samples) of PSUs that were created by SNZ for the delete-a-group jackknife estimation method as an approximation of the complete set of cluster-by-strata groupings that could be created if full information on PSUs and strata were available. SNZ formed 100 replication groups by randomly selecting PSUs from a list of PSUs, ordered by strata. In our implementation, each bootstrap replicate sample is formed through the random selection, with replacement, of 100 groups from the full set of 100. The members of each group may appear in the replicate sample once, more than once, or not at all. As in individual-level bootstrap sampling, the weights associated with each observation remain unchanged (weights are not adjusted for the presence or absence of a particular group of PSUs).

Replication testing within SNZ (Smith, 2001) indicated that around 100 replication groups captured enough of the sample variation to provide reasonable approximations of the variances that would be estimated using a full jackknife estimator. We make the premise, therefore, that the variation across replicate groups is also large enough to approximate the results of a full bootstrap estimator.

Sampling weights were used in the subsequent calculation of statistics, but not in the selection of the bootstrap replicate samples. Each bootstrap estimate is based on 100 replications.

Table A.1 compares the results that were obtained using several variations of the jackknife and the bootstrap methods. The 2003 Income Survey dataset was used for this comparative exercise. The statistics considered in the comparison were the mean and median of the total weekly income distribution, and the mean and median of the weekly earnings of wage and salary earners, for the following sub-populations: males, females, European/Pākehā, and Māori. This selection of income measures and sub-groups was intended to capture a variety of different dependent variable distributions, sample sizes, and sample design effects. To obtain the jackknife estimates for medians, we used the modified method proposed by Woodruff (1952), which is designed to give more robust results for standard errors on percentiles.

The two left-hand columns of Table A.1 give the statistics of interest and the standard errors that can be calculated on those statistics on the assumption that they are drawn from a sample without any design effects (i.e. using the conventional variance estimation formulas). The third and fourth columns of the table give standard errors estimated using the jackknife delete-a-group method, firstly without any adjustment for imputation and then with the imputed cases removed and the remaining sample re-weighted. Not surprisingly, reducing the underlying sample sizes raises the size of standard errors, particularly for Māori (reflecting the fact that the imputation rate is higher for Māori than for Pākeha).[16]

The fifth column gives standard errors estimated using a simple bootstrap procedure, using sample weights but without any adjustment for sample clustering. Column 6 gives standard errors obtained from the bootstrap method with the (simplified) approximation for sample clustering that is described above.

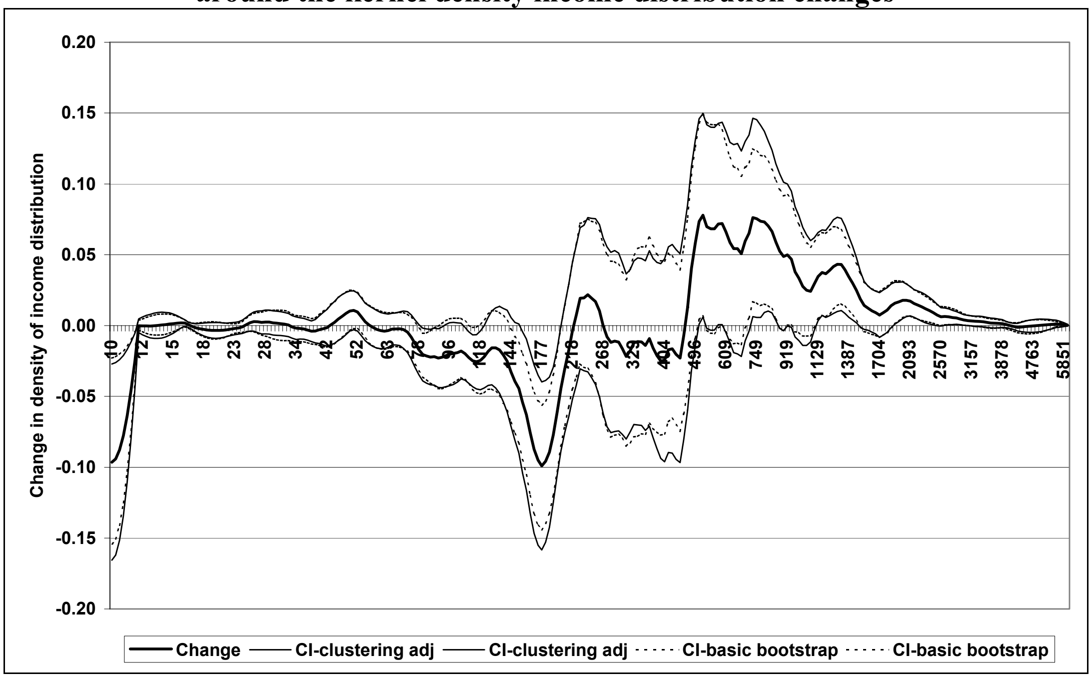
The results of the jackknife and bootstrap methods that take sample clustering into account (columns 4 and 6) are quite similar, although the bootstrap errors tend to be slightly larger. Both methods give standard errors for means that are around 1.5 times the simple random sample errors, although this ratio varies across estimates from 1.3 to 1.7.

The bootstrap method with the adjustment for clustering was used in the remainer of this study, simply because it is easier to apply than the jackknife method when the statistics of interest are percentiles.

[16] The method used in column 4 is closest to the method used by SNZ for IS standard error estimation.

Figure A.1 illustrates the impact that the choice of a bootstrap method has on the size of standard errors on the kernel density income distribution functions. It shows 95% confidence intervals around the base estimate of the total change in the Māori income distribution between 1997/98 to 2002/03 (as previously shown in Figure 9). The two sets of confidence intervals were estimated using the two alternative bootstrap methods described above, with and without an adjustment for the clustering of the sample. Although the confidence intervals are clearly larger when the more conservative method is implemented, the variation in method has relatively little impact on the regions of income for which the density changes fall within, versus outside the confidence intervals.

Figure A.1: Alternative bootstrap methods of estimating confidence intervals around the kernel density income distribution changes



References

Shao, Jun and Tu Dongsheng. 1995. *The Jackknife and Bootstrap*. Springer.

Smith, Harry. 2001. 'Investigation of the delete-a-group jackknife variance estimator for the HLFS,' Statistics New Zealand report, available on the SNZ website, www.stats.govt.nz.

Woodruff, R.S. (1952) 'Confidence intervals for medians and other position measures.' *Journal of the American Statistical Association,* 47, 635-646.

Table A.1: Comparison of standard error estimates using alternative estimation methods

	Base estimate (1)	SEs assuming a simple random sample, with weights (2)	Jackknife, using SNZ replicate weights (3)	Jackknife, using SNZ replicate weights & with imputed cases removed (4)	Bootstrap, using sample weights and with imputed obs removed (5)	Bootstrap, as in (5) & also modified for sample clustering (6)	Ratio jackknife (4) estimate to random sample SE (4/2)	Ratio bootstrap (5) estimate to random sample SE (5/2)	Ratio bootstrap (6) estimate to random sample SE (6/2)
Mean total weekly income for the total population									
Males	639.6	6.0	8.2	8.6	8.4	8.8	1.5	1.4	1.5
Females	386.9	3.3	4.7	4.3	4.7	5.1	1.3	1.4	1.5
European/Pakeha	544.9	4.1	5.6	6.0	5.7	6.6	1.4	1.4	1.6
Maori	439.1	7.2	7.5	9.9	10.4	9.6	1.4	1.4	1.3
Median total weekly income for the total population									
Males	522.6		6.2	7.0	9.5	9.6			
Females	288.8		1.9	1.9	2.3	2.0			
European/Pakeha	405.0		4.5	6.3	5.6	6.1			
Maori	371.3		8.3	10.0	10.4	8.4			
Mean total weekly earnings of w&s earners									
Males	780.4	6.9	9.6	10.4	9.2	10.8	1.5	1.3	1.6
Females	524.8	4.6	6.7	7.0	5.9	7.8	1.5	1.3	1.7
European/Pakeha	676.1	5.1	7.7	7.8	7.7	9.0	1.5	1.5	1.8
Maori	578.6	10.2	10.5	14.8	15.4	13.5	1.4	1.5	1.3
Median total weekly earnings of w&s earners									
Males	683.0		6.8	6.5	6.1	7.9			
Females	479.5		6.8	6.5	6.1	6.8			
European/Pakeha	600.0		3.4	5.0	5.6	6.3			
Maori	532.0		10.8	14.0	13.1	13.0			

Note: These estimates were obtained using the 2003 IS dataset and all persons aged 15 and over. The method illustrated in column (4) is closest to the method used by Statistics New Zealand for this survey. Column 5 is a simple bootstrap method using individual-level sampling with replacement. Column 6 incorporates modifications designed to capture the effects of sample clustering on sampling errors.

Appendix B: Kernel density estimation

The kernel density estimate $\hat{f}_h$ of an univariate distribution f, based on a random sample w_1, w_n with weights θ_1 , θ_n,, $\sum \theta_i = 1$, is

$$\hat{f}_h(w) = \sum_{i=1}^{n} \frac{\vartheta_i}{h} K\left[\frac{w - w_i}{h}\right]$$

where h is the bandwidth and $K(.)$ is the kernel function, which together regulate the relationship between the distance of w_i from w and the weight given to observation i in the estimation of the density at w.

In this paper, we use the Epanechnikov kernel function, a bandwidth of 0.08 for estimating Māori income distributions, and a bandwidth of 0.06 for estimating Pākehā/European income distributions. These bandwidths are slightly narrower than the 'optimal' bandwidths, which differ for the two groups because of the large differences in sample sizes. The 'optimal' bandwidths for the main income distributions considered in this paper, estimated using the method proposed by Silverman (1986), are as follows: 0.108 for all Pākehā; 0.067 for employed Pākehā; 0.135 for all Māori; and 0.087 for employed Māori.

In general, it is preferable to under-smooth than over-smooth as the latter may disguise localised peaks and troughs in the income distribution. This is the approach adopted in this paper.

The choice of bandwidth does affect the precision with which the income density function is estimated. In general, the smaller the bandwidth, the larger the sampling error on the estimate of the density function at a given point in the income distribution. Figures B.1 and B.2 illustrate the impact that a substantial (50%) increase in bandwidth size (from 0.08 to 0.12) has on the regularity of the kernel density income density function and the size of its 95% confidence intervals. They show the change between 1997/98 and 2002/03 in the income distribution of all working-aged Māori, estimated using bandwidths of 0.08 and 0.12 respectively. Figure B.1 corresponds to Figure 9 in the core of the paper. Although the income ranges for which the density changes are evaluated as ‘statistically significant’ are somewhat larger when a larger bandwidth is used, the difference is not large enough to materially change the substantive conclusions drawn.

References

Silverman, B. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Figure B.1: Confidence intervals on the change in the density of the log income distribution for working-aged Māori, using bandwidth of 0.08

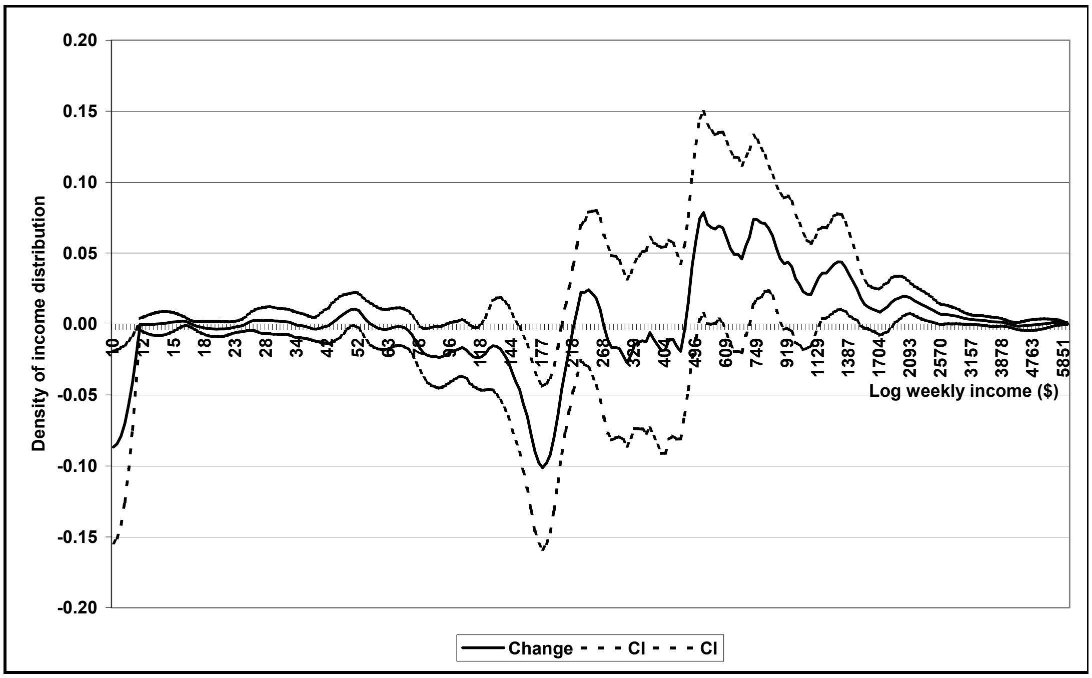
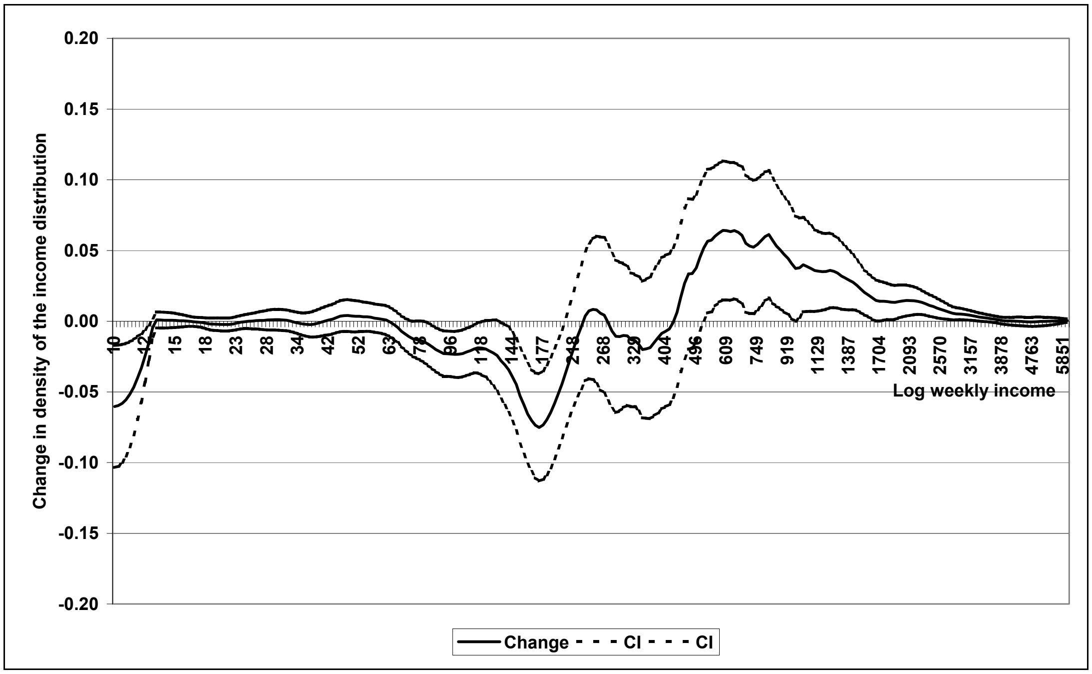


Figure B.2: Confidence intervals on the change in the density of the log income distribution for working-aged Māori, using bandwidth of 0.12



Motu Working Paper Series

04-11. Grimes, Arthur, "New Zealand: A Typical Australasian Economy?"

04-10. Viv B Hall and C. John McDermott, "Regional business cycles in New Zealand: Do they exist? What might drive them?"

04-09. Grimes, Arthur, Suzi Kerr and Andrew Aitken, "Bi-Directional Impacts of Economic, Social and Environmental changes and the New Zealand Housing Market".

04-08. Grimes, Arthur, Andrew Aitken, "What's the Beef with House Prices? Economic Shocks and Local Housing Markets".

04-07. McMillan, John, "quantifying Creative Destruction: Entrepreneurship and Productivity in New Zealand".

04-06. Maré, David C and Izi Sin, "Maori Incomes: Investigating Differences Between Iwi"

04-05. Kerr, Suzi, Emma Brunton and Ralph Chapman, "Policy to Encourage Carbon Sequestration in Plantation Forests".

04-04. Maré, David C, "What do Endogenous Growth Models Contribute?"

04-03. Kerr, Suzi, Joanna Hendy, Shuguang Liu and Alexander S.P. Pfaff, "Uncertainty and Carbon Policy Integrity".

04-02. Grimes, Arthur, Andrew Aitken and Suzi Kerr, "House Price Efficiency: Expectations, Sales, Symmetry".

04–01. Kerr, Suzi; Andrew Aitken and Arthur Grimes, "Land Taxes and Revenue Needs as Communities Grow and Decline: Evidence from New Zealand".

03-19. Maré, David C, "Ideas for Growth?".

03–18. Fabling, Richard and Arthur Grimes, "Insolvency and Economic Development:Regional Variation and Adjustment".

03–17. Kerr, Suzi; Susana Cardenas and Joanna Hendy, "Migration and the Environment in the Galapagos:An analysis of economic and policy incentives driving migration, potential impacts from migration control, and potential policies to reduce migration pressure".

03–16. Hyslop, Dean R. and David C. Maré, "Understanding New Zealand's Changing Income Distribution 1983–98: A Semiparametric Analysis".

03–15. Kerr, Suzi, "Indigenous Forests and Forest Sink Policy in New Zealand".

03–14. Hall, Viv and Angela Huang, "Would Adopting the US Dollar Have Led To Improved Inflation, Output and Trade Balances for New Zealand in the 1990s?"

03–13. Ballantyne, Suzie; Simon Chapple, David C. Maré and Jason Timmins, "Movement into and out of Child Poverty in New Zealand: Results from the Linked Income Supplement".

03–12. Kerr, Suzi, "Efficient Contracts for Carbon Credits from Reforestation Projects".

03–11. Lattimore, Ralph, "Long Run Trends in New Zealand Industry Assistance".

03–10. Grimes, Arthur, "Economic Growth and the Size & Structure of Government: Implications for New Zealand".

03–09. Grimes, Arthur; Suzi Kerr and Andrew Aitken, "Housing and Economic Adjustment".

03–07. Maré, David C. and Jason Timmins, "Moving to Jobs".

03–06. Kerr, Suzi; Shuguang Liu, Alexander S. P. Pfaff and R. Flint Hughes, "Carbon Dynamics and Land-Use Choices: Building a Regional-Scale Multidisciplinary Model".

03–05. Kerr, Suzi, "Motu, Excellence in Economic Research and the Challenges of 'Human Dimensions' Research".

03–04. Kerr, Suzi and Catherine Leining, "Joint Implementation in Climate Change Policy".

03–03. Gibson, John, "Do Lower Expected Wage Benefits Explain Ethnic Gaps in Job-Related Training? Evidence from New Zealand".

03–02. Kerr, Suzi; Richard G. Newell and James N. Sanchirico, "Evaluating the New Zealand Individual Transferable Quota Market for Fisheries Management".

03–01. Kerr, Suzi, "Allocating Risks in a Domestic Greenhouse Gas Trading System".

All papers are available online at http://www.motu.org.nz/motu_wp_series.htm