AGENDA
A JOURNAL OF POLICY ANALYSIS & REFORM

Volume 26, Number 1, 2019

Editor and Editorial Committee

ANALYSIS

Income inequality in New Zealand: Why conventional estimates are misleading
John Creedy and Norman Gemmell

When I'm 64: What do New Zealanders want in a retirement income policy?
Joey Au, Andrew Coleman and Trudy Sullivan

ARGUMENT

Government failure and success: A trans-Tasman comparison of two insulation subsidy schemes
Nicholas Preval, Jenny Ombler, Arthur Grimes, Michael Keall and Philippa Howden-Chapman
Editor

Dr William Coleman 
College of Business and Economics, ANU

Editorial Committee

Professor Harry Clarke 
University of Melbourne

Professor Max Corden 
University of Melbourne

Professor Anne Daly 
University of Canberra

Professor Kevin Davis 
University of Melbourne

Professor Henry Ergas 
University of Wollongong

Dr Stephen Grenville 
Lowy Institute, Sydney

Dr Arthur Grimes 
Motu Research, Wellington

Professor Kim Hawtrey 
Nimble Economics

Professor Doug Irwin 
Dartmouth College

Dr Gary Johns 
Australian Institute for Progress

Professor Tony Makin 
Griffith University

Dr Deborah Peterson 
Sapere Research Group

Professor David Vines 
Balliol College, Oxford

Dr Ruth Williams 
La Trobe University
ANALYSIS
Income inequality in New Zealand: Why conventional estimates are misleading

John Creedy and Norman Gemmell

Abstract

Considerable attention is now paid to establishing the extent of inequality in New Zealand and whether it has risen in recent years. This paper offers some insights into the inequality measures and interpretations that commonly feature in those debates. These typically relate annual Gini coefficients for various income definitions, or comparisons of income growth rates across income deciles. But cross-sectional data fail to take into account the longitudinal dimension of inequality, and this can lead to misinterpretations of inequality data. The paper shows that examining longitudinal income data for the same individuals over time strongly contradicts some apparent messages of cross-sectional evidence. For example, some recent cross-sectional inequality measures suggest that the incomes of initially low-income households grew at slower rates than those with initially higher incomes. This has been interpreted as the poorest earners being ‘left behind’. But recent longitudinal data, at least for individuals, reveals evidence of much faster-than-average growth among initially lower, compared to higher, income earners. Thus, ‘regression to the mean’ is a dominant feature of the longitudinal data.
Introduction

A great deal of attention is currently being paid to establishing the extent of inequality in New Zealand and whether it has risen in recent years. With the work of the Labour Government’s 2018 Tax Working Group well underway, and a proposed Welfare Working Group also to be established, the ability of tax and welfare policy to affect the inequality properties of the income distribution has also become a central political issue.

When summarising inequality, commentators generally recognise the need to specify the precise income measure, whether it is pre-tax or disposable income, or consumption expenditure. They also usually make the important distinction between individual and household or family incomes, where use of the latter recognises income sharing within households and differing needs of adults and children, along with possible economies of scale in consumption expenditure. Hence, results are often obtained where the ‘household income per adult equivalent person’ is treated as if it is obtained equally by each household member.\(^2\)

However, considerably less attention is usually paid to the time period, or accounting period, over which income is measured, yet this has important implications not only for orders of magnitude but for attitudes to measured inequality. This is easily illustrated using the simplified case of just two individuals (A and B) and two time periods (1 and 2), where 100 income units are available in total in each period (and discounting is ignored). Suppose A receives 80 units in period 1 and 20 units in period 2, while B receives 20 units in period 1 and 80 units in period 2. The only difference between the individuals is that they have different time patterns of income.

When measured over a single period there is substantial inequality (in each year the richer person obtains four times the income of the poorer person), but over the two periods they have equal total incomes of 100 units each. The attitude to inequality in this case is likely to differ substantially from a situation in which there is rigidity in the distribution and person A receives 80 units in each period. Of course, attitudes are also likely to depend significantly on the reasons for the relative income changes, and while some equalising mobility is likely to be regarded as ‘good’, reflecting opportunities for improvement, there could be ‘excessive’ mobility if this

---

\(^2\) A set of ‘adult equivalent scales’ is used by which each individual is assigned an ‘adult equivalent size’ (with a value of 1 for adult males). This assumption is not ‘innocent’, in that equal sharing is a strong assumption, made pragmatically in the face of a lack of information. Furthermore, the use of the individual as the ‘income unit’ in producing inequality measures can actually lead to a preference for inequality: that is, a transfer from a poorer to a richer household may be preferred if the richer household is larger and has substantial economies of scale in consumption. For an introduction to the issues involved in the choice of income unit, see Creedy (2017).
reflects considerable uncertainty. The choice of time period clearly depends, like that of the income and inequality measures used, on value judgements: there is no unambiguously ‘correct’ or ‘objective’ measure.5

The relationship between short- and long-period measures of inequality, while complex, clearly depends on the precise nature of relative income mobility. While the most common summary measures are reported for annual incomes, it is argued here that comparisons should be supplemented by information about such mobility. This requires the use of longitudinal data, whereby individuals or households can be tracked over a number of periods. Although such data are relatively scarce for New Zealand, failure to take into account the longitudinal and mobility dimensions of inequality can lead to misinterpretations of reported inequality measures.

For example, most of the inequality measures quoted in recent New Zealand debates relate to cross-sectional inequality, such as Gini coefficients for various income definitions across a number of years. When considering income changes over time, comparisons of income growth rates of deciles of the annual income distributions are made, using only cross-sectional data. Hence, this does not identify how far the same people remain towards the bottom or top of the income distribution, or the extent to which movement within the distribution occurs over time.

This paper examines some scarce longitudinal income data for individual New Zealand taxpayers and shows that, whereas inequality measures based on population cross-sections tend to suggest that the income of (initially) lower-income groups grew at similar or slower rates than those with initially higher incomes, this is strongly contradicted by longitudinal data. Indeed, there is conclusive evidence of faster-than-average growth for those with initially lower incomes. The phenomenon of ‘regression to the mean’ seems to be a dominant feature of the longitudinal data.4

First, section 2 briefly summarises the evidence usually provided about cross-sectional (annual) income inequality in New Zealand, and about differential income growth rates. Section 3 demonstrates how longitudinal income data, to measure relative income changes of individuals (rather than changes in deciles of annual income), provide valuable additional information that has typically been ignored in popular debates. This section makes use of Inland Revenue Department data which track a constant sample of individuals over a number of years. The data necessarily refer to individuals, which is the relevant unit used for income tax assessments.

3 The choice also depends on the context. For example, weekly incomes display considerable variability compared with longer-period incomes, yet a case may be made for a redistributive transfer system based on weekly incomes. Broader judgements about income inequality and the desired progressivity of an income tax structure may continue to be based on the use of a longer accounting period. For example, students may be poor in annual terms but subsequently rich in terms of lifetime incomes.

4 This phenomenon has also been observed to be strong in earlier studies for New Zealand: see, for example, Creedy (1996), Creedy et al. (2018).
Hence, they cannot reflect the way these are combined into families or households, or the extent to which they may be shared. Brief conclusions are in section 4. All dollar values discussed are in New Zealand dollars.

Cross-sectional inequality evidence

Contributors to New Zealand policy debates on income inequality invoke a variety of inequality measures for alternative definitions of income. Two readily available indicators are the taxable income distributions from Inland Revenue and the numerous distributional measures reported in Ministry of Social Development (MSD, 2017, and previous annual editions). Two examples are given below to illustrate the type of comparison given. Table 1 shows the percentage of taxpayers, taxable income and assessed tax paid by the top and bottom 10 per cent (approximately) of the tax-paying population, and similar values for the top and bottom 50 per cent.

### Table 1. Percentage shares: 2016 distribution of individual taxpayers in New Zealand (excluding zero taxable incomes)

<table>
<thead>
<tr>
<th>Year: 2016</th>
<th>Taxpayers (%)</th>
<th>Taxable income (%)</th>
<th>Assessed tax (%)</th>
<th>Income range ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom</td>
<td>9.7</td>
<td>0.7</td>
<td>0.4</td>
<td>1 to 7,000</td>
</tr>
<tr>
<td>Top</td>
<td>10.0</td>
<td>33.5</td>
<td>45.4</td>
<td>87,001 to 150,000+</td>
</tr>
<tr>
<td>Bottom</td>
<td>49.5</td>
<td>17.2</td>
<td>10.6</td>
<td>1 to 31,000</td>
</tr>
<tr>
<td>Top</td>
<td>50.5</td>
<td>82.8</td>
<td>89.4</td>
<td>31,001 to 150,000+</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations using Inland Revenue data: www.ird.govt.nz/aboutir/external-stats/revenue-refunds/income-distrib-individual-customers/income-distrib-individ-customers.html.

Table 1 reveals that the lowest 9.7 per cent of taxpayers (earning from $1 to $7,000 per year) received less than 1 per cent of all taxable income, while the top 10 per cent (earning over $87,000 in 2016) received around 34 per cent of total taxable income. However, those low-income taxpayers are on very low incomes. By comparison, the annual gross-of-tax New Zealand Superannuation (pension) in 2016 was around $23,000 for a single person. Hence the lowest 10 per cent of taxpayers are likely to be part-time earners, students or children, who are all subject to large transitory income changes over time; see Creedy, Enright, Gemmell and McNabb (2010) for a detailed decomposition of the lowest income taxpayers by type. This serves to

---

5 New Zealand, by virtue of having no tax-free zone, has a relatively large number of income taxpayers compared with Australia. It is hard to make comparisons because Australian income taxpayer data are generally reported for numbers of individual income tax returns, whilst only around 2.2 million of New Zealand’s 3.7 million individual income taxpayers are required to submit a tax return, or receive a ‘personal tax summary (PTS)’. Where a PTS is issued by Inland Revenue, this obviates the need to submit a tax return unless the taxpayer considers the PTS is incorrect. The tax liabilities of remaining income taxpayers in New Zealand are fully covered by the pay-as-you-earn system.
emphasise the limited value of focusing on the lowest income decile (or indeed the lowest two deciles, with incomes below $15,000 in 2016) since these are unlikely to provide a representative picture of individuals in income poverty.

Based on the same data source, Figure 1 shows how the share of total taxpayers with incomes of $0, $1 to $5,000, and $1 to $7,000 changed over the 2001–16 period for which comparable data are available. Figure 1 shows how the share of those on very low incomes has changed over the 2001–16 period. It can be seen that $0 earners represent about 3 to 4 per cent of all taxpayers, with a slight decline in recent years. The figure also shows the share of taxpayers earning less than $5,000 and $7,000 per year. This reveals a steady decline in both groups, which might be expected since these nominal income thresholds represent declining real incomes over the period. However, CPI data for the period shows that $5,000 in 2001 is approximately equal in real terms to $7,000 in 2016. Hence, in real terms, the 2016 value for ‘$1 to $7,000’ can be compared with the 2001 value for ‘$1 to $5,000’. This indicates that the share of taxpayers with those low real incomes fell from around 11.2 per cent in 2001 to 9.4 per cent in 2016. That is, the proportion of the lowest (real) income taxpayers has fallen over the 15 years.

Figure 1. Percentage of all income taxpayers on zero or low incomes: New Zealand 2001 to 2016
Source: Authors’ calculations using Inland Revenue data: www.ird.govt.nz/aboutir/external-stats/revenue-refunds/income-distrib-individual-customers/income-distrib-individ-customers.html.
These data are of course highly selective, focusing on small segments of the tax-paying population, and are based on a series of annual cross-sections of the income distribution. More comprehensive coverage of household incomes, rather than individual taxpayers, but still based on annual cross-sectional data, is provided by MSD in their regular publication on household incomes in New Zealand; see, for example, MSD (2016, 2017). Since households differ by size and composition, as mentioned in section 1, MSD use total household income per adult equivalent person as their income measure, and use the individual as unit of analysis. Two of their examples are shown in Figures 2 and 3.

Figure 2 shows deciles of the distribution of annual household incomes after housing costs, for 1982 to 2016 (using data from MSD (2017, p. 76, Table D6)). Their commentary states:

> from a longer-term perspective: in HES 2016, household incomes at the top of the bottom decile were no better than they were in the 1980s. This is the only decile for which this is the case, though for P20 the gain is small (MSD, 2017, p. 75).

More generally, these data have been interpreted incorrectly as evidence of the lower-income deciles ‘falling behind’ the top deciles in real income growth, or even that real incomes are falling absolutely for lower-income households.

These messages appear to be reinforced by evidence on the income growth rates by decile of the income distribution, such as those shown in Figure 3, compiled from the same data used for Figure 2, for a ‘post global financial crisis period’ (2009 to 2016). This shows larger income growth rates for those in initially higher-income deciles compared to lower deciles. That is, richer households seem to be moving away from poorer households in terms of their real income levels, over the period.

---

6 Overall patterns look similar when based on incomes before housing costs (BHC), though there is more evidence of higher real incomes for the lowest deciles at the end of the period compared to the 1980s; see MSD (2017, p. 75).

7 For example, examining data for income growth over a longer period, 1982 to 2015, MSD (2016, p. 65) conclude that ‘all income groups gained in real terms, with the highest income group gaining much more than the rest, and the lowest income group gaining the least. The different growth rates show that income inequality is higher in HES 2015 than in 1982’. However, they later point out that this cannot capture income changes for ‘the same individuals over time’ (MSD, 2016, p. 65). Similarly, Rashbrooke (2013) misleadingly claims, ‘in the last thirty years, incomes for our richest New Zealanders has more than doubled, while those of the poorest have barely increased – and in fact are lower now than they were in the 1980s if you take their increased housing costs into account; see www.inequality.org.nz/understand/narrative/. These statements give the misleading impression that they relate to income growth of those with initially lower and higher incomes yet they are not based on longitudinal data.

8 The growth rates here refer to income growth rates for the household at each decile: 1 (at the 10th percentile), 2 (at the 20th percentile), and so on, of the income distribution, not average growth rates for households between relevant deciles; for example, between the 10th and 20th percentiles.
Figure 2. Real equivalised household incomes after housing costs at decile boundaries, 1982–2016
Source: Authors’ calculations using MSD (2017, p. 76, Table D6).

Figure 3. Real equivalised household incomes: Changes at each decile
Source: Authors’ calculations using MSD (2017, p. 76, Table D6).
However, such an interpretation is based on a misunderstanding of the difference between ‘income growth rates for different individuals with the same low initial incomes’ and ‘income growth for the same individuals with initially low incomes’. The former is observed from a series of cross-sectional ‘snapshots’ such as those mentioned previously, whereas the latter requires longitudinal data for the same individuals over time. Longitudinal evidence for New Zealand is examined in the next section.9

Inequality evidence from longitudinal data

To examine longitudinal income data comparable to the cross-sectional evidence for households presented in section 2, it would be necessary to track the same households over time. Unfortunately, data to allow this for New Zealand are currently limited to the Survey of Family Income and Employment (SoFIE) data, which are only available for the 2002 to 2010 period, for a relatively small sample.10 With the increasing scope and availability to researchers of matched household-level data in the Integrated Data Infrastructure at Statistics New Zealand, such longitudinal household income analyses may soon be possible.

However, it should be recognised that the use of households is considerably complicated by the fact that their size and composition changes over time, as a result of births, deaths, family division and formation, and so on. The use of individual data in this section avoids those complications, but at the cost that changes in individuals’ income over time may not necessarily indicate equivalent changes in household incomes. Nevertheless, observed differences between cross-sectional and longitudinal patterns of income change for individuals might be expected to provide some guidance on the potential for similar differences at the household level.

9 MSD (2016, p. 65) points out this distinction by stating that, In interpreting the time series analysis that is based on the HES data (as above), it is important to understand that the HES provides repeat cross-sectional data with different people interviewed each survey. The HES does not follow the same individuals across time. Some individuals do stay in roughly the same income band for many years, some move up and some move down. The degree of income mobility in New Zealand is discussed in Section K using longitudinal data from Statistics New Zealand’s Survey of Family, Income and Employment (SoFIE).

Unfortunately, SoFIE data cover only a limited number of years: 2002–09, and many commentators continue to ignore the above warning.

10 See Carter and Gunasekara (2012) for some SoFIE-based mobility evidence. There are around 18,000 individuals in each of the seven waves of their SoFIE samples. The number of individuals in the Inland Revenue data used in section 3 exceed 30,000.
With access to Inland Revenue’s confidentialised longitudinal individual taxpayer data, Creedy and Gemmell (2017, 2018) examined the growth rates of taxpayer incomes across the individual (as opposed to household) income distribution for three five-year periods: 1998 to 2002, 2002 to 2006 and 2006 to 2010. This section reports on some of those longitudinal results and assesses them against directly comparable cross-sectional evidence.

The data used in this section are for a 2 per cent random sample of individual New Zealand Inland Revenue personal income taxpayers. Using data for 2002, 2006 and 2010, results for two separate panels are reported for 2002 to 2006 and 2006 to 2010, each (5-year) panel containing incomes for both years for the same taxpayers. This yielded useable samples of 31,355 and 32,970 individuals respectively. To avoid the exercise being contaminated by taxpayers with very low incomes (such as small part-time earnings of children, or small capital incomes of non-earners), individuals with annual incomes less than $1,000 were omitted from the sample.

Appendix Table A1 provides details of income levels in each sample – averages within income ventiles (20 equal-sized divisions) and at equivalent percentiles (0.05, 0.10, 0.15 etc.). As might be expected using individual data, even eliminating incomes below $1,000, there are still substantial numbers of taxpayers on low incomes. For example, the sample 25th percentile income in 2006 was $16,924, around the level of the New Zealand state pension (NZ Superannuation) for a married person.11 Individual taxpayers in the lowest income ventiles in this dataset might therefore be expected in general to be members of households that include other income sources.

To identify longitudinal aspects, in each case individuals were ranked by their initial year incomes (2002 or 2006), with all of the diagrams below showing percentiles of the income distribution in the relevant initial year (2002 or 2006) on the horizontal axis. For cross-sectional comparisons, individuals were ranked in each of the two years, regardless of their rank in the previous or subsequent year. That is, they represent income distributions of different rankings of individuals that ignores their prior or subsequent status.

Figures 4 and 5, covering the periods 2002 to 2006 and 2006 to 2010 respectively, capture both longitudinal and cross-sectional aspects of inequality in the form of several Lorenz curves. In each panel, two cross-sectional Lorenz curves are shown where individuals are ranked by their incomes in the initial and final year: 2002 and 2006 in Figure 4; and 2006 and 2010 in Figure 5.

11 Most taxpayers receiving only NZ Superannuation are typically in the 2nd and 3rd deciles; see www.ird.govt.nz/aboutir/external-stats/revenue-refunds/income-distrib-individual-customers/income-distrib-individ-customers.html. The longitudinal sample percentile incomes in Appendix Table A1 are somewhat different from those reported in Table 1 previously, which refers to all New Zealand income taxpayers.
Figure 4. Lorenz curves: New Zealand taxpayers 2002 to 2006
Source: Authors’ calculations using unpublished Inland Revenue confidentialised individual taxpayer data (a 2 per cent random sample made available to the authors).

Figure 5. Lorenz curves: New Zealand taxpayers 2006 to 2010
Source: Authors’ calculations using unpublished Inland Revenue confidentialised individual taxpayer data (a 2 per cent random sample made available to the authors).
In both figures, the two Lorenz curves can be seen to be almost indistinguishable from each other, indicating little difference in a Gini-based cross-sectional measure of inequality for each year. For 2010, the Lorenz curve mostly lies slightly outside the 2006 equivalent, though some crossing of the two curves can also be seen. Two important messages arise from those charts.

First, it is known that increasing the length of the accounting period tends to reduce inequality measures such as the Gini coefficient, for example as incomes are measured over 1, 2, 5, 10 years and so on.\(^{12}\) This aspect is evident in Figure 4 and Figure 5, where incomes in both years for each individual are added (and the average ranked by initial year incomes). While this does not include incomes for all years over the two five-year periods, it does demonstrate the effect of extending the accounting period.

It can be seen that this Lorenz curve for two years of income data lies wholly inside the two equivalent curves for 2002 and 2006 (Figure 4) or 2006 or 2010 (Figure 5). That is, longer-term incomes display noticeably less inequality than either of the annual cross-sections. These curves confirm that substantial reductions in inequality, as the income period is lengthened, are consistent with the relatively unchanged cross-sectional patterns that are also observed in Figures 4 and 5.

Secondly, in Figure 5, when longitudinal dimensions are introduced explicitly by ordering income in both 2006 and 2010 for each individual by their 2006 incomes, a further distinct reduction in inequality is apparent. That is, the ‘2010 longitudinal (ranked by 2006)’ curve lies everywhere closer to the line of equality than any of the other curves in the chart. Similar patterns can be observed in Figure 4 for 2002 to 2006. The Gini coefficient is the area between the Lorenz curve and the line of equality, expressed as a fraction of the total area below the line of equality. Hence, it follows that for both these periods the ‘longitudinal-based Gini’ must be smaller than the cross-sectional based Gini.

An alternative, and in some ways more insightful, means of comparing the different inequality outcomes captured by cross-sectional and longitudinal dimensions of the data, is to examine growth rates over the two periods, 2002 to 2006 and 2006 to 2010, by different groups within the (initial) income distribution. These are shown by ventiles of the income distribution in Figures 6 and 7, for 2002 to 2006 and 2006 to 2010 respectively, which allow for a more fine-grained comparison than income deciles. The cross-sectional growth rates by ventile are obtained by ranking individual incomes from lowest to highest separately for each year. Growth rates are then estimated for the incomes of the lowest to the highest individual, who will generally \textit{not} be the same person in both years.

\(^{12}\) See, for example, Laws (2014) and Creedy et al. (2018) for New Zealand evidence.
These growth rates are then averaged within ventiles; with over 30,000 individuals in total there are in excess of 1,500 in each ventile. These are shown by the hatched bars in Figures 6 and 7, with values on the horizontal axis representing the ventile, 0.5, 0.10, ..., 0.95, 1.0. Longitudinal growth rates, on the other hand, relate to the income growth of the same individual with ventiles based on the initial income ranks (in 2002 or 2006).

Figure 6. Income growth rates: New Zealand taxpayers 2002 to 2006
Source: Authors’ calculations using unpublished Inland Revenue confidentialised individual taxpayer data (a 2 per cent random sample made available to the authors).

Figure 7. Income growth rates: New Zealand taxpayers 2006 to 2010
Source: Authors’ calculations using unpublished Inland Revenue confidentialised individual taxpayer data (a 2 per cent random sample made available to the authors).
It can be seen in Figure 6, for 2002 to 2006, that growth rates based on cross-sectional data are similar across ventiles, with most generally close to the mean for the whole sample, and there is no obvious pattern towards systematically higher or lower growth rates across the ventiles. By contrast, the (solid) longitudinal-based growth rates demonstrate that income growth over the five years was systematically higher for low-income, compared to high-income, ventiles, with clear evidence of substantial regression towards the mean across the ventiles.

For example, on average individuals who were initially in the lowest 5 per cent of incomes in 2002 experienced much faster income growth over the period 2002 to 2006 than individuals in the highest ventiles. Indeed, for those in above-median ventiles, income growth was negative on average, especially for the top 5 per cent of taxpayers. A similar pattern is observed in Figure 7 for 2006 to 2010, but here the cross-sectional income growth rates demonstrate some tendency for higher ventiles to experience faster growth than lower ventiles (at least for those from around 0.2 upwards). Longitudinal-based ventiles, however, continue to reveal that when the same individuals are observed, the initially lowest income taxpayers experienced the fastest subsequent income growth.

In both Figures 6 and 7 a strong pattern is the negative association of income growth rates with ventile income levels across the whole distribution; that is, it is not simply the lowest and highest ventiles that display this tendency. Hence, although the income data for individuals used here include some very low incomes (that would be expected to be much less prevalent with household-level data), it seems that differences in the cross-sectional/longitudinal growth rate patterns are not simply due to income volatility experienced by those with low incomes in one year.

Figures 6 and 7 therefore highlight that if inequality of income growth experience is assessed by reference to cross-sectional growth rates across ventiles or deciles, it can lead to the impression that the poorest are being ‘left behind’, as those on initially higher incomes experience faster growth. However, this obscures the evidence that if income growth rates are based on the initial income distribution, this conclusion is not supported; indeed, it is radically reversed. Some initially higher-income individuals move down the income distribution over time, and vice versa, such that when comparing the incomes of the ‘old lowest’ to the ‘new lowest’, observed income growth rates are similar.\(^\text{13}\)

---

\(^{13}\) Using data from the Household, Income and Labour Dynamics in Australia (HILDA) survey, Azpitarte (2014) reports a similar contrast between income growth rates obtained using longitudinal and cross-sectional data.
Given these ‘regression towards the mean’ patterns, it might be expected that income inequality measured over several years would be less than when measured over a single year. In fact, this motivated Shorrocks (1978) to propose a measure of income mobility or ‘rigidity’ based on comparing an index of longer-term income inequality to single-year income inequality.\(^{14}\)

This phenomenon has been observed for a similar longitudinal sample of New Zealand income taxpayers by Creedy, Gemmell and Laws (2018), covering a 19-year period from 1994. These are illustrated in Figure 8, which shows Gini inequality indices for taxable incomes over those years. In each case the profiles show the Gini index calculated using incomes aggregated over 1, 2, 3, … 19 years, beginning in 1994 or 2001.\(^{15}\) Profiles are shown for all taxpayers in the sample and for ‘working-age’ taxpayers only (those aged 24–64 throughout the period) – since retirement can be associated with a substantial drop in income around age 65.

All four profiles show that the inclusion of incomes for an additional year consistently results in a reduction in the Gini coefficient, with the exception of year 7 for the ‘1994 start’ profile; that is, when the 2001 income data are included. This reflects the impact of the increase in the top rate of income tax from 33 per cent to 39 per cent in 2001 which was associated with (especially) top incomes being diverted from the personal income tax schedule and possibly increased income sharing among partnered individuals. The figure therefore confirms, as expected, that allowing for taxable incomes received over a longer period generates a lower inequality outcome than is observed for each annual income distribution on its own. For example, the ratio of the Gini for 19 years of income to the Gini for the first year (1994) is 0.811 (all ages) and 0.855 (working age).\(^{16}\)

---

\(^{14}\) The Shorrocks (1978) measure of ‘rigidity’, \(R\), involves constructing the ratio of an index of the inequality of \(T\)-year incomes to an equivalent weighted average index of income inequality in the \(T\) single years.

\(^{15}\) Data for a longitudinal panel from 2001 are also shown since this period follows the income tax reforms in 2000 which can be seen to have a modest (upward) impact on the Gini profiles which include that year; see Creedy et al. (2018) for further discussion.

\(^{16}\) Though these ratios are not exactly the equivalent of the Shorrocks (1978) measure, the values are similar to estimates of the Shorrocks R measure reported for Australia, based on 16-year income averages, in Wilkins and Lass (2018). They find \(R=0.836\) (all ages) and \(R=0.813\) (age 18–29 in 2001); see Wilkins and Lass (2018, p. 34). For New Zealand, estimates of Shorrocks’s R calculated over 10 years from 2003, yield \(R=0.896\) (all ages) and \(R=0.897\) (age 24–64).
Figure 8. Gini indices for multi-year incomes: New Zealand taxpayers from 1994 and 2001

Source: Authors’ calculations using unpublished Inland Revenue confidentialised individual taxpayer data (a 2 per cent random sample made available to the authors).
Conclusions

The clear conclusion from this examination is that a great deal of care is required when interpreting evidence on inequality based on income growth rates across deciles (or other divisions) of the income distribution. Since contributors to inequality and poverty debates often begin by asking how increases in incomes or welfare of initially poorer groups compare with those initially better off, the relevant evidence is longitudinal, not cross-sectional.

On the other hand, if concern is with the poorest in any or all periods for which data are being examined, then care needs to be taken to recognise the extent to which people move between income bands over time. In this case, estimating inequality using Gini or similar indices that are measured over several years provides a clearer picture of longer-term inequality as distinct from that which is observed from annual data which include a mixture of persistent and transitory components.

An important caveat to the analysis reported here, as stressed in section 3, is that the comparisons of cross-sectional and longitudinal measures of income inequality are based here on income of individuals. Much of the debate about the level of, or changes in, inequality tends to relate to household inequality, both in New Zealand and Australia. Without access to suitable longitudinal household income data for New Zealand (at least outside of 2002–10), the present analysis has focused on individual taxpayer data. This suggests some caution is warranted with respect to conclusions regarding differences that longitudinal approaches to measuring household inequality and mobility may identify.

However, the evidence here of substantial regression towards the mean across most longitudinal income ventiles, while cross-section–based equivalents display no such tendency, is suggestive that conclusions for household income inequality or mobility could be misleading if based on cross-section data for households. Hopefully, future research will be able to clarify how far these patterns across individual-level incomes are a feature of household-level equivalents.

References


Appendix: New Zealand taxable income distributions 2002 and 2006

Table A1 provides summary measures of the distribution of taxable income for individuals in 2002 and 2006. These measures are obtained using the 2 per cent samples discussed in section 2.

Table A1. Percentile income levels and ventile average incomes: New Zealand taxpayers ($)

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Income at percentile</th>
<th>Average ventile income</th>
<th>Income at percentile</th>
<th>Average ventile income</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5,641</td>
<td>3,307</td>
<td>7,177</td>
<td>4,214</td>
</tr>
<tr>
<td>10</td>
<td>8,750</td>
<td>7,455</td>
<td>10,371</td>
<td>9,076</td>
</tr>
<tr>
<td>15</td>
<td>10,471</td>
<td>9,605</td>
<td>12,946</td>
<td>11,458</td>
</tr>
<tr>
<td>20</td>
<td>12,344</td>
<td>11,565</td>
<td>14,774</td>
<td>13,601</td>
</tr>
<tr>
<td>25</td>
<td>14,420</td>
<td>13,452</td>
<td>16,924</td>
<td>15,614</td>
</tr>
<tr>
<td>30</td>
<td>16,437</td>
<td>15,386</td>
<td>20,453</td>
<td>18,615</td>
</tr>
<tr>
<td>35</td>
<td>19,437</td>
<td>17,938</td>
<td>24,272</td>
<td>22,357</td>
</tr>
<tr>
<td>40</td>
<td>22,510</td>
<td>20,953</td>
<td>27,864</td>
<td>26,070</td>
</tr>
<tr>
<td>45</td>
<td>25,492</td>
<td>24,000</td>
<td>31,328</td>
<td>29,611</td>
</tr>
<tr>
<td>50</td>
<td>28,468</td>
<td>26,957</td>
<td>34,689</td>
<td>32,998</td>
</tr>
<tr>
<td>55</td>
<td>31,276</td>
<td>29,884</td>
<td>37,693</td>
<td>36,193</td>
</tr>
<tr>
<td>60</td>
<td>34,097</td>
<td>32,710</td>
<td>40,550</td>
<td>39,044</td>
</tr>
<tr>
<td>65</td>
<td>36,942</td>
<td>35,486</td>
<td>44,084</td>
<td>42,306</td>
</tr>
<tr>
<td>70</td>
<td>39,753</td>
<td>38,304</td>
<td>47,951</td>
<td>45,993</td>
</tr>
<tr>
<td>75</td>
<td>43,293</td>
<td>41,470</td>
<td>52,517</td>
<td>50,161</td>
</tr>
<tr>
<td>80</td>
<td>47,876</td>
<td>45,532</td>
<td>57,923</td>
<td>55,079</td>
</tr>
<tr>
<td>85</td>
<td>53,849</td>
<td>50,715</td>
<td>63,541</td>
<td>60,506</td>
</tr>
<tr>
<td>90</td>
<td>61,012</td>
<td>57,661</td>
<td>73,412</td>
<td>68,099</td>
</tr>
<tr>
<td>95</td>
<td>79,146</td>
<td>68,355</td>
<td>94,504</td>
<td>82,503</td>
</tr>
<tr>
<td>100*</td>
<td>150,033</td>
<td>130,329</td>
<td>186,437</td>
<td>160,030</td>
</tr>
<tr>
<td>Overall mean</td>
<td>34,038</td>
<td></td>
<td>41,156</td>
<td></td>
</tr>
</tbody>
</table>

Note: * This row shows income at the 99th percentile rather than the top income (100th percentile). However, average ventile income in this row includes all individuals from the 95th to the 100th percentile.

Source: Authors’ calculations using unpublished Inland Revenue confidentialised individual taxpayer data (a 2 per cent random sample made available to the authors).
When I’m 64: What do New Zealanders want in a retirement income policy?

Joey Au, Andrew Coleman and Trudy Sullivan

Abstract

It is difficult to choose policies when people have diverse preferences over outcomes and many alternative policy settings are available. To do this well, policymakers must understand underlying preferences and rank policies according to these preferences. In this paper, we use multi-criteria decision analysis techniques to understand the relative attractiveness of retirement policy reforms in New Zealand. Using a nationally representative sample, we estimate individual preferences over seven aspects of retirement policy, characterise the diversity of these preferences, and rank three different policy options. We find that a policy which raises taxes to prefund the government retirement income scheme would be supported by a majority of people of all ages and income groups, and would be much more popular than a policy that
raises the age of eligibility. The results suggest multi-criteria decision analysis has considerable potential to help policymakers develop policies that are aligned with people’s preferences.

1 Introduction

Governments use a variety of methods including public opinion polling and focus group discussions to elicit their constituents’ preferences (Lees-Marshalment, 2015). These methods typically require a trade-off between the complexity of the issues that are investigated and the statistical validity of the resultant findings. In response to these trade-offs, researchers in the health, environment and energy fields sometimes use multi-criteria decision analysis techniques to obtain systematic evidence about people’s preferences (Devlin and Sussex, 2011; Gamper and Turcanu, 2007; Mendoza and Martins, 2006). While these techniques are designed to analyse the relative importance of different aspects of complex problems (Arrow and Raynaud, 1986; Belton and Stewart, 2002), they have yet to be used to study major public finance problems.

This paper reports the findings of an online multi-criteria decision survey conducted by the New Zealand Treasury to investigate public preferences about retirement income policy reforms. The paper (i) investigates the relative importance of seven aspects of retirement income policies to a representative sample of New Zealanders, (ii) documents the diversity of their policy preferences in a systematically quantifiable manner, and (iii) ranks three different retirement income policies from an individual welfare perspective.

Several other retirement income surveys have asked respondents to make explicit trade-offs between two aspects of a policy, for example whether they would be prepared to raise taxes to provide larger retirement incomes (e.g. Boeri, Boersch-Supan and Tabellini (2002) for Italy and Germany; Van Els et al. (2004) for the Netherlands; or Fourati and O’Donoghue (2009) for Ireland). This study improves their survey methodologies in three ways. First, rather than surveying respondents about their attitudes towards complex policy packages, respondents are asked to make pairwise comparisons over particular features of policy packages, decisions that are easier to make. Secondly, the technology enables us to estimate a complete ranking of the relative importance of each criterion for each person, which previous studies have been unable to do. For example, we estimate the relative importance of the pension amount and the age of eligibility, the pension amount and future tax rates, and the pension amount and the means-testing regime, not just one of these combinations. Thirdly, the technique allows us to estimate each respondent’s preferences over large numbers of policy packages comprised of
different combinations of simple policy features rather than the small number of policies included in traditional surveys. Thus the technique can be used to inform the development of new policies.

The survey approach proved to be very successful and suggests multi-criteria analysis is a promising technique to improve the ways governments design policies. One of our main findings, for instance, is that a substantial majority of respondents would opt for an increase in current taxes if this could prevent even larger tax increases on future generations. We also find that there is widespread opposition to means-testing, and that there are strongly divergent preferences over the appropriate eligibility age for New Zealand Superannuation. These findings enable us to conclude that a policy combination that raises the eligibility age for New Zealand Superannuation and reduces future tax increases would be supported by few and opposed by many. However, a policy that more aggressively prefunds the government pension by immediately raising taxes and investing surplus funds would be supported by a majority of people of all ages and income levels.

2 The retirement income survey

2.1 Multi-criteria decision analysis

Multi-criteria decision analysis has been developed to assist people make complex choices over outcomes that involve multiple criteria or dimensions in an explicit, consistent and transparent way. This paper uses the PAPRIKA (Potentially All Pairwise Rankings of all Possible Alternatives) method implemented through 1000Minds software to estimate a respondent’s preference ranking over a set of criteria (Hansen and Ombler, 2008; Ombler and Hansen, 2012). Respondents are presented with a series of hypothetical choices in an online survey, each of which involves scenarios that combine a highly-ranked category from one criterion and a lowly-ranked category from another, so that each choice requires a trade-off to be made. Figure 1 is an example of a trade-off question from the survey. Respondents choose the combination of criteria they prefer from the two alternative scenarios: the one on the left retains the age of eligibility at 65 but requires current taxes to increase by 2 per cent; the one on the right keeps current taxes the same, but raises the age of eligibility to 67. A respondent chooses his or her preferred combination, or indicates that they are indifferent between the two scenarios. Once the selection is made, the respondent is presented with another hypothetical scenario using categories from two randomly selected criteria. The process is repeated until the algorithm has enough information to estimate a complete preference ranking over the criteria.
Which of these two scenarios do you prefer the most?  
(given they’re identical in all other aspects)

You should not be concerned about whether either of these policies can be delivered.  
We simply want to know if you prefer one scenario more than the other.

Age when NZ Superannuation starts  
65 years (current policy)  
Extra taxes to be paid now?  
everyone pays 2% more taxes  
(EXAMPLE: $20 more each week if earning $50,000)

or

Age when NZ Superannuation starts  
67 years (2 years later)  
Extra taxes to be paid now?  
no extra taxes

this scenario

they are equal

Figure 1. Example of a trade-off question  
Source: Authors’ survey using 1000Minds software.

Each criterion in the survey is represented by categories that are ranked from lowest to highest according to the benefits they provide a person. For example, if the categories for the ‘age of eligibility’ criterion are ‘65 years’ and ‘67 years’, the category ‘65 years’ would be ranked higher than ‘67 years’ as a person receives a pension for more years, and can retire earlier. The PAPRIKA method reduces the number of questions that respondents answer by automatically excluding dominated pairwise comparisons and by imposing transitivity to implicitly answer other questions.

The PAPRIKA algorithm assumes each respondent’s preferences can be represented by a cardinal utility function defined over the survey criteria, and it estimates a set of utility weights for each criteria-category that is consistent with the way they answer the survey.\footnote{The working paper version of this paper includes a simplified description of the algorithm for those wanting more information. The algorithm is fully described in Hansen and Ombler (2008).} For example, if a person prefers a pension of $360 per week from age 65 over a pension of $390 per week from age 67, the sum of the weights for ‘$360 per week’ and ‘65 years’ exceeds the sum of the weights for ‘$390 per week’ and ‘67 years’. Since the weights are inversely proportional to the rank of each criteria when there are only two categories for each criterion, in this paper we usually present the results in terms of the rank of each criterion from 1 (important) to 7 (unimportant), noting that the most important criterion has the highest weight.  
(Note that all dollar amounts discussed throughout are in New Zealand dollars.)
The weights are used to estimate how individuals rank policies that differ in terms of the outcomes they deliver. The calculation is done in two steps. First, the effect of the policy on each of the criteria is estimated. The policy might increase the age of eligibility to 66 years and increase the amount of the pension to $375 per week, for example. Secondly, the effect of the policy on each individual’s utility is estimated by adding up the individual’s weights for each outcome, using linear interpolation where necessary. For example, if a policy raised the age of eligibility to 66, we would use the average of the weights for ‘65 years’ and ‘67 years’. The policy with the highest utility score is deemed to be the individual’s preferred policy.

2.2 New Zealand’s government retirement income schemes

Government retirement income schemes are classified three ways. Tier 1 schemes provide a retirement income funded from general taxation to eligible people irrespective of the amount they contribute during their working-age years. These incomes can be the same for all people (universal) or they can be means-tested. Tier 2 or contributory schemes provide retirement incomes that depend on the amount that people contribute to the scheme during their working-age years. Tier 3 schemes encourage voluntary retirement saving by offering subsidies or lower taxes. New Zealand is one of two OECD countries without a tier 2 scheme. It has a universal tier 1 scheme, New Zealand Superannuation, that is largely funded on a pay-as-you-go basis from general taxation, and a subsidised tier 3 scheme, KiwiSaver, introduced in 2007.

The survey was designed to find out the population’s preferences over several policy reforms under discussion at the time of the survey. These potential reforms included: (i) raising the age of eligibility for New Zealand Superannuation; (ii) prefunding New Zealand Superannuation by raising current taxes and investing the proceeds; (iii) introducing a means-test for New Zealand Superannuation; and (iv) introducing a tier 2 compulsory saving scheme.

2.3 The survey criteria

The survey criteria listed in Table 1 were chosen after a process involving an extensive review of the retirement income policy literature, the results of a trial survey conducted on a non-representative group of public servants, and discussions with several focus groups. The number of categories was kept to a minimum to reduce the length of the survey. Each criterion had two category levels that were chosen so that the differences between categories were broadly similar in dollar terms. The baseline

---

3 The focus groups included students, retirees, women, Māori, Pacific people, disabled people, retirement policy experts and representatives from Grey Power, an advocacy group promoting the welfare of older citizens.
categories for the age of eligibility (age 65) and the amount of the pension ($360 per week) reflect the parameters of the New Zealand Superannuation scheme in 2014. The second category for the age of eligibility, 67 years, was chosen as it is an age that was often mentioned in contemporaneous public debate. As two years of retirement income is approximately equal to $30,000, or $30 per week over the average length of time someone receives a pension, the second category of the pension amount was chosen to be $390 per week, so that the two criteria could be meaningfully compared. The baseline categories for the current and future tax criteria are the taxes that would be needed to fund New Zealand Superannuation now and in the future if there were no changes to its current structure. The second category reflects the size of the tax increase that would be necessary to support a $30 per week increase in the size of the pension, which was calculated to be approximately 2 per cent of personal income. For symmetry we chose to vary current and future taxes by the same amount. Since New Zealand Superannuation is universally provided rather than means-tested, the baseline category was ‘No means-testing’. We chose the second category of the means-testing criterion to make the revenue raised by the means-test similar to the revenue saved by raising the age of eligibility.

Table 1. The survey criteria

<table>
<thead>
<tr>
<th>The criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amount of NZ Superannuation everyone receives</strong></td>
<td><strong>Age when NZ Superannuation starts</strong></td>
</tr>
<tr>
<td>• $360 a week (current level)</td>
<td>• 67 years (2 years later)</td>
</tr>
<tr>
<td>• increases by $30 a week to $390</td>
<td>• 65 years (current policy)</td>
</tr>
<tr>
<td><strong>Extra taxes to be paid now?</strong></td>
<td><strong>Extra taxes the next generation (i.e. not you) has to pay</strong></td>
</tr>
<tr>
<td>• everyone pays 2% more taxes (Example: $20 more each week if earning $50,000)</td>
<td>• 5% more taxes (Example: $50 more each week if earning $50,000)</td>
</tr>
<tr>
<td>• no extra taxes</td>
<td>• 3% more taxes (Example: $30 more each week if earning $50,000)</td>
</tr>
<tr>
<td><strong>Will everyone receive the same amount of NZ Superannuation?</strong></td>
<td><strong>The amount of your personal savings to spend or invest when you retire</strong></td>
</tr>
<tr>
<td>• No, people with retirement savings greater than $200,000 have their NZ Superannuation reduced by $60 per week</td>
<td>• 2 years of your average annual income (don’t worry how you get this amount)</td>
</tr>
<tr>
<td>• Yes, everyone gets the same NZ Superannuation</td>
<td>• 3 years of your average annual income (don’t worry how you get this amount)</td>
</tr>
<tr>
<td><strong>Savings flexibility</strong></td>
<td></td>
</tr>
<tr>
<td>• it is compulsory to save 5% of your income each week (Example: $50 put aside each week if earning $50,000)</td>
<td><strong>you can save when and how you like</strong></td>
</tr>
<tr>
<td>• you can save when and how you like</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ survey criteria distributed using 1000Minds software.
One issue that underlies the whole survey is framing. It is well known that the framing of questions can have an enormous effect on survey responses. Some authors such as Bartels (2003) argue that framing effects may be sufficiently crucial to the design of a survey that they fundamentally undermine the use of all surveys as a source of useful information. We – and all other authors who conduct surveys – are not so extreme in our views. Nonetheless, it is possible that the relative ranking of the responses in part reflects the way the questions were framed, and that the answers might have been different if they were framed differently.

3 Survey results and discussion

3.1 The sample

A professional surveying firm, Colmar Brunton, provided the web panel sample, which was conducted in 2014. Respondents typically answered 12 survey questions, and took 5 to 10 minutes to answer the survey. To assuage concern that the respondents may not have understood the surveying technique, the respondents were required to repeat two of the comparison questions at the end of the survey. These comparison questions included one of the most preferred and one of the least preferred criteria, making it easy for someone doing the survey in good faith to answer, while discriminating against respondents who may not have understood the questions or who answered the questions in a random fashion. Over 1,300 people were surveyed, of whom 1,066 or 80 per cent answered both repeated questions consistently.\(^4\) This success rate provides evidence that the survey procedure was well understood. Respondents who did not answer both additional questions consistently were excluded from the sample. People who answered the survey very quickly were also excluded, as their responses were often inconsistent.

Table 2 provides information about selected economic and demographic characteristics of the survey respondents. The results have been reweighted to take account of the difference between the socio-demographic characteristics of survey respondents and the socio-demographic characteristics of the total population, using weights provided by Colmar Brunton. The reweighting has little effect on the results as the sample is broadly representative of the New Zealand public. The sample of 1,066 is similar in size to the Eurobarometer surveys analysed by

\(^4\) We also tested the PAPRIKA algorithm transitivity assumption. Consider the two tax criteria ‘current taxes’ and ‘future taxes’. Some respondents compared these two criteria directly, while for others the relative importance of the two criteria is inferred indirectly by applying the transitivity principle to their responses to other questions. The average ranks of the ‘direct’ and ‘indirect’ groups should be the same. Of the 63 per cent of respondents who compared the two criteria directly, 66 per cent ranked low future taxes as more important than low current taxes. Of the remaining 37 per cent of respondents, 63 per cent ranked future taxes as more important than current taxes. The similarity of these two numbers in this and other examples suggests the transitivity assumption holds adequately in this survey.
Lynch and Myrskylä (2009), and while it is smaller than some other retirement income surveys it is large enough to generate confidence interval estimates that can discriminate between alternative hypotheses as the intervals are small relative to the scales we are using.\(^5\) In addition to demographic information, respondents were asked how confident they are that they will have enough money to live comfortably in retirement. Four options were offered: ‘Not confident at all’, ‘Not too confident’, ‘Somewhat confident’ and ‘Very confident’. The responses to this question were one of the strongest predictors of an individual’s preferences. Of the sample group, 9 per cent said they were ‘Not confident at all’, while 16 per cent said they were ‘Very confident’.

Table 2. Selected economic and demographic characteristics of the survey respondents

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>Household income</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>46%</td>
<td>26%</td>
<td>32%</td>
</tr>
<tr>
<td></td>
<td>&lt;35</td>
<td>&lt;$50,000</td>
<td>Degree</td>
</tr>
<tr>
<td>Female</td>
<td>54%</td>
<td>59%</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td>35–64</td>
<td>$50,000-$100,000</td>
<td>Post-sec. school</td>
</tr>
<tr>
<td></td>
<td>65+</td>
<td>$100,000+</td>
<td>High school</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Employment status</td>
<td>Geographical spread</td>
<td>Confidence in retirement</td>
</tr>
<tr>
<td>European</td>
<td>76%</td>
<td>Full-time</td>
<td>Auckland</td>
</tr>
<tr>
<td>Māori</td>
<td>12%</td>
<td>Part-time</td>
<td>Other North Island</td>
</tr>
<tr>
<td>Pacific</td>
<td>6%</td>
<td>Retired</td>
<td>South Island</td>
</tr>
<tr>
<td>Other</td>
<td>6%</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>KiwiSaver member?</td>
<td>Children</td>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>67%</td>
<td>Yes</td>
<td>Married/civil union</td>
</tr>
<tr>
<td>No</td>
<td>32%</td>
<td>No</td>
<td>De facto</td>
</tr>
<tr>
<td>Don’t know</td>
<td>1%</td>
<td></td>
<td>Single</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations derived from the survey responses.

---

\(^5\) The Eurobarometer surveys surveyed between 1,000 and 1,100 respondents in 15 European countries, and 600 respondents in Luxembourg. Other retirement income surveys have taken place in Ireland, with 326 respondents (Fourati and O’Donoghue, 2009); in the Netherlands, with 1,200 households (Van Els et al., 2004); in the United States, with 1,400 respondents (the General Social Survey; Ponza et al., 1988); in Australia, with 1,600 respondents in three different years (Evans and Kelley (2004)); and in Italy and Germany, with 2,500 respondents in each country (Boeri et al., 2002).
3.2 Methodological approach

A respondent’s survey response can be represented by estimated utility weights or by a vector that lists the rank he or she gives to each criterion, for example $x_i = (6 \ 5 \ 2 \ 1 \ 4 \ 7 \ 3)$. The equivalence of the ranking and utility weighting representations means non-parametric methods can be used to analyse the results. If there is a set of $m$ respondents, $X = \{x_1, x_2, \ldots, x_m\}$, one measure of their average preferences is the mean preference vector $\bar{x}$:

$$\bar{x} = \frac{1}{m} \sum_{j=1}^{m} x_j$$  \hfill (1)

If there are two populations, the difference in their mean preference vectors is an estimate of the differences in their average preferences. We use the Wilcoxon-Mann-Whitney statistic to test the hypothesis that members of the two groups have the same distribution of preferences over a particular criterion. If the test statistic has a high absolute value, the hypothesis that the two groups have the same distribution of preferences over this criterion can be rejected. We also use the Li and Schucany (1975) statistic to test the hypothesis that the mean preference vectors are the same for all of the criteria, not just each single criterion.

We measure the diversity of preferences of a group by calculating the mean Spearman rank correlation coefficient between all possible pairs of people in the group. To calculate this, we first calculate the Spearman rank correlation between the vectors of two members of a group, $x$ and $y$: if $n$ is the number of criteria, the correlation coefficient is

$$\rho_{xy} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^{n} (x_i - y_i)^2$$  \hfill (2)

The mean Spearman rank correlation is calculated by calculating the mean Spearman rank correlation for all $m(m+1)/2$ possible pairs of respondents. This has a value between negative one and one. If a group has uniformly distributed preferences, meaning that each of the vectors is equally likely, it would have a mean Spearman value equal to zero.

---

6 This section is based on Mardin (1995).
7 We also calculated a different set of metrics that correspond to a group’s median responses. As these results are qualitatively similar in all respects, we have chosen to only discuss results in terms of mean responses. The other results are available in the working paper version of the paper.
3.3 The average level and dispersion of preference ranks

Table 3 shows the mean preference vector for the sample, plus the fraction of the population who ranked each criterion either highest or lowest. (Rank ‘1’ means the criterion is important; rank ‘7’ means it is unimportant.) There are three key results.

Table 3. Average retirement income preferences in New Zealand (N=1,066)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>1 Pension amount</th>
<th>2 Age 65/67</th>
<th>3 Current taxes</th>
<th>4 Future taxes</th>
<th>5 Means tests</th>
<th>6 Wealth amount</th>
<th>7 Flexible savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall importance</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mean rank</td>
<td>4.09</td>
<td>3.92</td>
<td>4.15</td>
<td>3.41</td>
<td>3.15</td>
<td>4.27</td>
<td>5.02</td>
</tr>
<tr>
<td>% rank 1</td>
<td>7.8%</td>
<td>21.2%</td>
<td>5.9%</td>
<td>12.4%</td>
<td>41.7%</td>
<td>5.4%</td>
<td>9.8%</td>
</tr>
<tr>
<td>% rank 7</td>
<td>10.7%</td>
<td>15.1%</td>
<td>5.3%</td>
<td>2.6%</td>
<td>14.3%</td>
<td>8.8%</td>
<td>38.1%</td>
</tr>
</tbody>
</table>

Mean Spearman correlation 0.080

Source: Authors’ calculations derived from the survey responses.

(i) The distribution of preferences. New Zealanders have very diverse preferences about retirement income policies. Nevertheless, while the mean Spearman rank correlation is only 0.08, the hypothesis that New Zealanders have uniformly distributed preferences can be rejected at the 5 per cent significance level.

(ii) The relative importance of the seven criteria. The mean preference ranks for the different criteria range from a minimum of 3.15 to a maximum of 5.02. The two highest-ranked criteria concern (i) universality/means-testing and (ii) future tax rates. On average, respondents expressed a strong preference for universal rather than means-tested pensions, and were opposed to policies that result in steep increases in taxes on future generations. The lowest-ranked criterion is the flexible saving/compulsory saving criterion: few people thought saving flexibility provided many advantages relative to a compulsory saving scheme. The other four criteria had mean ranks very close to 4, the mean value when preferences are uniformly distributed. Three of these criteria – current tax levels, the amount of the pension, and the amount of wealth people have in retirement – were of moderate importance to most people. The fourth criterion, the age of eligibility, had a bimodal distribution.

(iii) The distribution of rank preferences. Figures 2A–2F show the entire distributions of the preference rank vectors. Figure 2A shows that the ‘universality/means-testing’ criterion is the most important criterion to the largest number of people. Forty-two percent of respondents rank universality rather than means-testing as the most important feature of retirement income policy, and an additional 20 per cent rank it as the second or third most important.
Figure 2B shows the distribution of preference ranks for the two tax criteria. Both are ‘hump-shaped’: they are moderately important but neither the most important nor the least important criteria for most people. The figure shows most people are more opposed to future tax increases than current tax increases: indeed, 65 per cent of all respondents gave (low) ‘future taxes’ a higher weight than ‘current taxes’, while only 30 per cent ranked them the other way around. (This split held for all population subgroups.) These responses strongly suggest that there is widespread opposition to policies that impose high costs on future generations.

Figures 2C and 2D show the distribution of preference ranks for two other hump-shaped criteria: the benefit of higher pensions, and the benefit of higher retirement savings. These criteria are moderately important to most people with respondents being almost equally divided as to whether they preferred higher pensions or higher retirement savings.

Figure 2E shows the distribution of preference ranks for the flexible saving/compulsory saving criterion. It has the opposite shape to Figure 2A: there are relatively few people who think saving flexibility is very important, and many who think it is not important. Overall, 38 per cent of respondents indicated that saving flexibility was the least important of all seven criteria, and only 19 per cent indicated it was one of the two most important criteria. This result suggests there would be little opposition to a compulsory saving scheme if it raised the amount of wealth available at retirement, possibly because many people already save this amount.

Lastly, Figure 2F shows the preference ranks for the age of eligibility criterion. It is the only criterion with a bimodal response. Thirty-seven per cent of the respondents indicate it is very important to keep the age of eligibility at 65 (1st or 2nd ranking), and 32 per cent of respondents indicate it is unimportant (6th or 7th ranking). The criterion is important to people from low-income households, to New Zealanders of Pacific ethnicity, and to those who are not confident about their retirement prospects, but unimportant to people over 65, to New Zealanders of European ethnicity, and to people who are confident about their retirement prospects.
Figure 2A. Universality/No means-testing

Figure 2B. Current and future taxes

Figure 2C. Pension amount

Figure 2D. Wealth amount

Figure 2E. Saving flexibility/compulsion

Figure 2F. Age of eligibility

Figure 2. Distribution of preference ranks for the whole survey

Source: Authors’ calculations derived from the survey responses. Rank 1=highest rank/most important criterion.
3.4 Results for population subgroups

It is natural to ask whether different population subgroups have different preferences. The short answer is that they do, but that these differences, while statistically significant, tend to be small. With one exception (the subgroup of Pacific people, discussed further below), there were few criteria where the mean preference ranks for population subgroups defined in terms of observable characteristics such as age, gender, household income, education or ethnicity differed by more than 0.5 ranks on a scale of 1–7. Indeed, the largest differences between subgroups occurred for groups that self-identified in terms of their expected comfort in retirement rather than for groups that could be identified in terms of measurable characteristics.

The analysis of population subgroups is conducted in two ways. First, we divided the population into subgroups and compared the mean preference ranks for a particular subgroup with all people not in that group: for example, people aged 65 or over versus people aged less than 65. In each case we calculated the mean rank of each criterion for the two subgroups, and used a Wilcoxon-Mann-Whitney test to test whether the distributions were the same. \(^8\) Table 4 shows the mean differences for each criterion for various subgroups. These subgroup comparisons do not condition on other factors that may vary across subgroups, and thus do not estimate the marginal effect of a socio-demographic factor on preference ranks. To do this we estimate a fractional multinomial logit model (Table 5), using the entire set of socio-demographic variables as independent variables. This model takes into account the loss of one degree of freedom that occurs when objects are ranked. We estimate the model using the utility function representation in which the rank vectors are converted into a set of normalised weights. As each criterion only has two categories, the weights are equal to the rank divided by 28, and sum to one. These weights are simultaneously regressed against dummy variables corresponding to each of the socio-demographic variable categories. The regression coefficients indicate how the weight of each criterion depends at the margin on each socio-demographic variable.

\(^8\) We also calculated the Li-Schucany test statistic of the hypothesis that the two subgroups have the same mean vector of preferences across all criteria: this was rejected at the 1 per cent level for all groups, and is not reported.
Table 4. Mean difference by criteria for selected population subgroups

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>1 Pension amount</th>
<th>2 Age 65/67</th>
<th>3 Current taxes</th>
<th>4 Future taxes</th>
<th>5 Means tests</th>
<th>6 Wealth amount</th>
<th>7 Flexible savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>All people</td>
<td>1,066</td>
<td>4.09</td>
<td>3.92</td>
<td>4.15</td>
<td>3.41</td>
<td>3.15</td>
<td>4.23</td>
<td>5.02</td>
</tr>
<tr>
<td><strong>Demographic characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>510</td>
<td>−0.23*</td>
<td>0.21</td>
<td>−0.07</td>
<td>0.14</td>
<td>−0.13</td>
<td>−0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>Female</td>
<td>556</td>
<td>0.23*</td>
<td>−0.21</td>
<td>0.07</td>
<td>−0.14</td>
<td>0.13</td>
<td>0.19</td>
<td>−0.26</td>
</tr>
<tr>
<td>Single</td>
<td>253</td>
<td>−0.10</td>
<td>0.05</td>
<td>−0.10</td>
<td>−0.07</td>
<td>0.29</td>
<td>−0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>Has children</td>
<td>624</td>
<td>0.13</td>
<td>−0.23</td>
<td>0.02</td>
<td>0.10</td>
<td>0.13</td>
<td>0.08</td>
<td>−0.23</td>
</tr>
<tr>
<td>Age &lt;35</td>
<td>308</td>
<td>0.32**</td>
<td>−0.05</td>
<td>−0.43**</td>
<td>−0.20</td>
<td>0.24</td>
<td>0.19</td>
<td>−0.07</td>
</tr>
<tr>
<td>Age 35–64</td>
<td>556</td>
<td>−0.06</td>
<td>−0.32*</td>
<td>0.12</td>
<td>0.07</td>
<td>−0.04</td>
<td>0.10</td>
<td>0.13</td>
</tr>
<tr>
<td>Age 65+</td>
<td>202</td>
<td>−0.34*</td>
<td>0.58**</td>
<td>0.38**</td>
<td>0.15</td>
<td>−0.25*</td>
<td>−0.42**</td>
<td>−0.11</td>
</tr>
<tr>
<td><strong>Region and ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auckland</td>
<td>353</td>
<td>−0.05</td>
<td>0.00</td>
<td>−0.18</td>
<td>0.01</td>
<td>−0.02</td>
<td>−0.04</td>
<td>0.27*</td>
</tr>
<tr>
<td>European</td>
<td>783</td>
<td>−0.16</td>
<td>0.39**</td>
<td>0.42**</td>
<td>−0.07</td>
<td>−0.42**</td>
<td>−0.28**</td>
<td>0.12</td>
</tr>
<tr>
<td>Māori</td>
<td>160</td>
<td>0.06</td>
<td>−0.20</td>
<td>0.07</td>
<td>−0.18</td>
<td>0.65**</td>
<td>0.05</td>
<td>−0.44*</td>
</tr>
<tr>
<td>Pacific</td>
<td>70</td>
<td>0.52**</td>
<td>−0.60**</td>
<td>−0.50**</td>
<td>−0.51**</td>
<td>0.80**</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>Asian</td>
<td>128</td>
<td>0.06</td>
<td>−0.29</td>
<td>−0.69**</td>
<td>0.36*</td>
<td>0.30</td>
<td>0.37**</td>
<td>−0.13</td>
</tr>
<tr>
<td><strong>Highest education and employment status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary school</td>
<td>340</td>
<td>−0.03</td>
<td>−0.13</td>
<td>0.19</td>
<td>0.19</td>
<td>−0.15</td>
<td>0.31**</td>
<td>−0.39*</td>
</tr>
<tr>
<td>Degree</td>
<td>466</td>
<td>0.00</td>
<td>0.30*</td>
<td>−0.22*</td>
<td>−0.16</td>
<td>0.11</td>
<td>−0.17</td>
<td>0.14</td>
</tr>
<tr>
<td>Full-time job</td>
<td>558</td>
<td>0.10</td>
<td>−0.02</td>
<td>−0.18</td>
<td>−0.04</td>
<td>−0.28</td>
<td>0.11</td>
<td>0.30*</td>
</tr>
<tr>
<td><strong>Household income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;$50,000</td>
<td>368</td>
<td>−0.21</td>
<td>−0.30*</td>
<td>0.06</td>
<td>0.27*</td>
<td>0.42**</td>
<td>0.08</td>
<td>−0.31**</td>
</tr>
<tr>
<td>$50,000–$100,000</td>
<td>426</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>−0.25*</td>
<td>−0.02</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>&gt;$100,000</td>
<td>272</td>
<td>0.17</td>
<td>0.29</td>
<td>−0.12</td>
<td>0.00</td>
<td>−0.48**</td>
<td>−0.18</td>
<td>0.32*</td>
</tr>
<tr>
<td><strong>Confidence about having enough money to live comfortably in retirement, and KiwiSaver membership</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not confident</td>
<td>86</td>
<td>−0.58**</td>
<td>−0.67**</td>
<td>0.11</td>
<td>0.39*</td>
<td>1.14**</td>
<td>−0.15</td>
<td>−0.25</td>
</tr>
<tr>
<td>Very confident</td>
<td>183</td>
<td>0.43**</td>
<td>0.64**</td>
<td>−0.37**</td>
<td>−0.17</td>
<td>−0.43*</td>
<td>0.30*</td>
<td>−0.41</td>
</tr>
<tr>
<td>KiwiSaver</td>
<td>701</td>
<td>0.03</td>
<td>−0.22</td>
<td>−0.13</td>
<td>−0.22</td>
<td>−0.30</td>
<td>−0.02</td>
<td>0.85**</td>
</tr>
<tr>
<td>No KiwiSaver</td>
<td>365</td>
<td>−0.03</td>
<td>0.22</td>
<td>0.13</td>
<td>0.22</td>
<td>0.30</td>
<td>0.02</td>
<td>−0.85**</td>
</tr>
</tbody>
</table>

Notes: Each statistic is the difference in the mean preference rank for the identified subgroup with all people not in that subgroup. A negative number means the mean rank is lower (more important) for the subgroup. A*(**) indicates the hypothesis that the two groups have the same distribution of preferences can be rejected at the 5 per cent (1 per cent) significance level, using a Wilcoxon-Mann-Whitney non-parametric test.

Source: Authors' calculations derived from the survey responses.
Table 5. Fractional multinomial logit estimates of the effect of socio-demographic variables on the relative ranking of different policy criteria

<table>
<thead>
<tr>
<th></th>
<th>1 Pension amount</th>
<th>2 Age 65/67</th>
<th>3 Current taxes</th>
<th>4 Future taxes</th>
<th>5 Universality/means tests</th>
<th>6 Wealth amount</th>
<th>7 Flexible savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper North Island excl. Auckland (ref: Auckland)</td>
<td>0.004 (0.006)</td>
<td>0.002 (0.007)</td>
<td>0.006 (0.005)</td>
<td>0.002 (0.005)</td>
<td>-0.008 (0.007)</td>
<td>0.004 (0.005)</td>
<td>-0.009 (0.007)</td>
</tr>
<tr>
<td>Lower North Island (ref: Auckland)</td>
<td>0.006 (0.005)</td>
<td>0.002 (0.006)</td>
<td>-0.001 (0.005)</td>
<td>-0.006 (0.005)</td>
<td>-0.002 (0.007)</td>
<td>-0.003 (0.005)</td>
<td>0.000 (0.006)</td>
</tr>
<tr>
<td>South Island (ref: Auckland)</td>
<td>-0.006 (0.005)</td>
<td>-0.009 (0.006)</td>
<td>0.003 (0.005)</td>
<td>0.006 (0.005)</td>
<td>0.008 (0.007)</td>
<td>0.002 (0.005)</td>
<td>-0.003 (0.006)</td>
</tr>
<tr>
<td>Male (ref: Female)</td>
<td>-0.007 (0.004)</td>
<td>0.007 (0.006)</td>
<td>-0.003 (0.004)</td>
<td>0.004 (0.005)</td>
<td>0.000 (0.005)</td>
<td>-0.005 (0.004)</td>
<td>0.004 (0.005)</td>
</tr>
<tr>
<td>35–64 years (ref: &lt;35 years)</td>
<td>-0.013 (0.005)*</td>
<td>-0.005 (0.006)</td>
<td>0.012 (0.005)*</td>
<td>0.007 (0.005)</td>
<td>-0.006 (0.007)</td>
<td>-0.008 (0.005)</td>
<td>0.002 (0.006)</td>
</tr>
<tr>
<td>65+ years (ref: &lt;35 years)</td>
<td>-0.033 (0.009)*</td>
<td>0.030 (0.012)*</td>
<td>0.028 (0.009)*</td>
<td>0.003 (0.009)</td>
<td>-0.006 (0.012)</td>
<td>-0.036 (0.008)*</td>
<td>0.016 (0.011)</td>
</tr>
<tr>
<td>NZ European and Māori (ref: NZ European only)</td>
<td>0.010 (0.009)</td>
<td>0.009 (0.011)</td>
<td>-0.004 (0.007)</td>
<td>-0.013 (0.007)</td>
<td>0.019 (0.011)</td>
<td>-0.015 (0.008)</td>
<td>-0.007 (0.011)</td>
</tr>
<tr>
<td>Māori only (ref: NZ European only)</td>
<td>0.005 (0.011)</td>
<td>-0.011 (0.012)</td>
<td>0.003 (0.008)</td>
<td>0.002 (0.012)</td>
<td>0.012 (0.008)</td>
<td>-0.013 (0.012)</td>
<td></td>
</tr>
<tr>
<td>Chinese only (ref: NZ European only)</td>
<td>-0.002 (0.01)</td>
<td>-0.003 (0.011)</td>
<td>-0.020 (0.009)*</td>
<td>0.006 (0.009)</td>
<td>0.003 (0.011)</td>
<td>0.011 (0.009)</td>
<td>0.005 (0.01)</td>
</tr>
<tr>
<td>Pacific only (ref: NZ European only)</td>
<td>0.015 (0.011)</td>
<td>0.006 (0.013)</td>
<td>-0.025 (0.01)*</td>
<td>-0.011 (0.009)</td>
<td>0.007 (0.013)</td>
<td>0.003 (0.01)</td>
<td>0.004 (0.013)</td>
</tr>
<tr>
<td>Other (ref: NZ European only)</td>
<td>0.004 (0.006)</td>
<td>-0.027 (0.007)*</td>
<td>-0.009 (0.006)</td>
<td>0.009 (0.006)</td>
<td>0.018 (0.008)*</td>
<td>0.005 (0.006)</td>
<td>0.000 (0.007)</td>
</tr>
<tr>
<td>Other post-secondary school qualifications (ref: School qualification or less)</td>
<td>-0.002 (0.005)</td>
<td>0.003 (0.007)</td>
<td>-0.001 (0.005)</td>
<td>-0.006 (0.005)</td>
<td>0.003 (0.007)</td>
<td>-0.013 (0.005)*</td>
<td>0.015 (0.006)*</td>
</tr>
<tr>
<td>University degree or equivalent (ref: School qualification or less)</td>
<td>-0.001 (0.005)</td>
<td>0.013 (0.006)*</td>
<td>-0.002 (0.004)</td>
<td>-0.008 (0.004)</td>
<td>0.010 (0.006)</td>
<td>-0.015 (0.005)*</td>
<td>0.002 (0.006)</td>
</tr>
<tr>
<td>Not working (ref: Working)</td>
<td>0.001 (0.006)</td>
<td>0.005 (0.007)</td>
<td>0.001 (0.005)</td>
<td>0.002 (0.005)</td>
<td>0.003 (0.007)</td>
<td>-0.011 (0.005)*</td>
<td>-0.001 (0.007)</td>
</tr>
<tr>
<td>Retired (ref: Working)</td>
<td>0.010 (0.01)</td>
<td>-0.006 (0.01)</td>
<td>-0.013 (0.008)</td>
<td>0.003 (0.008)</td>
<td>-0.010 (0.011)</td>
<td>0.006 (0.008)</td>
<td>0.010 (0.011)</td>
</tr>
<tr>
<td>$30,001–$50,000 (ref: $30,000 or less)</td>
<td>0.001 (0.008)</td>
<td>0.009 (0.01)</td>
<td>-0.007 (0.006)</td>
<td>-0.002 (0.006)</td>
<td>0.000 (0.008)</td>
<td>0.006 (0.007)</td>
<td>0.003 (0.009)</td>
</tr>
<tr>
<td>$50,001–$70,000 (ref: $30,000 or less)</td>
<td>0.000 (0.008)</td>
<td>0.020 (0.01)*</td>
<td>0.000 (0.006)</td>
<td>-0.009 (0.006)</td>
<td>-0.010 (0.008)</td>
<td>-0.009 (0.007)</td>
<td>0.007 (0.009)</td>
</tr>
<tr>
<td>$70,001–$100,000 (ref: $30,000 or less)</td>
<td>0.002 (0.008)</td>
<td>0.022 (0.01)*</td>
<td>-0.003 (0.007)</td>
<td>-0.009 (0.006)</td>
<td>-0.016 (0.008)*</td>
<td>-0.014 (0.006)*</td>
<td>0.019 (0.009)*</td>
</tr>
</tbody>
</table>
The dependent variable is the criterion weight equal to the relative rank/28. A negative coefficient means that people with the particular socio-demographic variable consider the criterion to be important.

Source: Authors’ calculations derived from the survey responses.

The results of the two approaches are broadly consistent. In the fractional multinomial logit model, many socio-demographic variables have a statistically significant effect on the mean rank of one or more criteria, but the effects tend to be small, less than one rank. For example, the estimated coefficient for people 65 and over on the ‘pension size’ criterion in Table 5 is –0.033, indicating that a large pension is more important to people aged 65 and over than people aged under 35. On average they give it a rank 0.9 (=0.033*28) lower (i.e. more important). Typically, when a socio-demographic variable is statistically significant in the fractional multinomial logit model, that social-demographic subgroup has a mean rank vector that differs from the rest of the population by a statistically significant amount.

The results can be analysed in terms of the criteria that are important to particular socio-demographic groups (i.e. an analysis of the rows of Tables 4 and 5), or in terms of the socio-demographic characteristics that are associated with people who have strong views about each of the seven criteria (i.e. an analysis of the columns of Tables 4 and 5). Considering the rows first, three distinctive features associated with different socio-demographic factors stand out. First, people aged 65 and over have different preferences than people aged less than 65: they think it is more important to have a higher pension and more wealth in retirement; they are more opposed to means-testing and less opposed to increases in current taxes; and they are less concerned with keeping the age of eligibility at 65. While these effects are statistically significant, they are relatively small, with average ranks differing by less than 0.6 and generally much less.

Secondly, people living in low-income households have a stronger preference for keeping the age of eligibility at 65 than other groups; they are also more opposed to compulsion but less concerned about means-testing or future tax rates. The same preferences are also shown by people who are not confident they will have enough

<table>
<thead>
<tr>
<th></th>
<th>1 Pension amount</th>
<th>2 Age 65/67</th>
<th>3 Current taxes</th>
<th>4 Future taxes</th>
<th>5 Universality/means tests</th>
<th>6 Wealth amount</th>
<th>7 Flexible savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$100,001 or more (ref: $30,000 or less)</td>
<td>0.003 (0.008)</td>
<td>0.025 (0.011)*</td>
<td>–0.004 (0.007)</td>
<td>–0.003 (0.007)</td>
<td>–0.023 (0.008)*</td>
<td>–0.019 (0.007)*</td>
<td>0.022 (0.01)*</td>
</tr>
<tr>
<td>De facto relationship (ref: Married/civil union)</td>
<td>–0.005 (0.006)</td>
<td>0.003 (0.007)</td>
<td>–0.002 (0.005)</td>
<td>0.004 (0.005)</td>
<td>0.001 (0.007)</td>
<td>–0.001 (0.006)</td>
<td>0.000 (0.007)</td>
</tr>
<tr>
<td>Single (ref: Married/civil union)</td>
<td>–0.004 (0.006)</td>
<td>0.011 (0.007)</td>
<td>–0.002 (0.005)</td>
<td>–0.002 (0.005)</td>
<td>0.002 (0.007)</td>
<td>–0.007 (0.005)</td>
<td>0.004 (0.006)</td>
</tr>
<tr>
<td>Have children (ref: No children)</td>
<td>0.010 (0.005)*</td>
<td>–0.004 (0.006)</td>
<td>–0.002 (0.004)</td>
<td>–0.001 (0.004)</td>
<td>–0.001 (0.006)</td>
<td>0.005 (0.004)</td>
<td>–0.008 (0.005)</td>
</tr>
<tr>
<td>Somewhat confident (ref: Very confident)</td>
<td>–0.008 (0.005)</td>
<td>–0.010 (0.007)</td>
<td>0.008 (0.006)</td>
<td>0.004 (0.005)</td>
<td>0.011 (0.008)</td>
<td>–0.013 (0.005)*</td>
<td>0.008 (0.007)</td>
</tr>
</tbody>
</table>
money to live comfortably in retirement, but they are more strongly held. The latter group also expresses a stronger preference for means-testing and would like the size of the pension increased.

Thirdly, the results in Table 4 suggest there are differences in the preferences of different ethnic groups. New Zealanders with non-European ethnicity tend to be more concerned with keeping the age of eligibility at 65, are more opposed to increases in current taxes, and are more supportive of means-tested pensions than New Zealanders with European ethnicity. The differences are most marked for Pacific people. However, it appears that these differences mainly reflect the different age and income characteristics of non-European New Zealanders, as ethnicity is not an important factor in the fractional multinomial logit regressions. Once age and income are taken into account in the fractional multinomial regressions, the results showing that Pacific people have relatively strong preferences for higher pensions, lower taxes, an earlier eligibility age and means-testing no longer hold.

The results can be restated by highlighting the socio-demographic factors that are associated with each of the survey criteria (i.e. the columns of Table 4 and 5). There are three significant results.

First, an increase in the age of eligibility tends to be supported by higher-income people, those of European ethnicity, those who are more confident about having enough money to live comfortably in retirement, and people 65 and over. Secondly, opposition to means-testing is higher among higher-income people, those of European ethnicity, those who are more confident about having a comfortable retirement, and people aged 65 and over. Thirdly, saving flexibility rather than compulsion is most strongly preferred by low-income people, those with fewest educational qualifications and — most strongly — by those who are not members of the KiwiSaver scheme. The former two groups are the people most likely to need to reduce consumption if there were a compulsory saving scheme, and thus those most likely to be inconvenienced by a compulsory saving scheme.

While these differences in the preferences of different socio-demographic groups are statistically significant, they are not typically large. For example, while universality is more important to people over 65 than under 65, it is the single most important criterion to both groups, and the mean rank for this criterion for those over 65 is only 0.25 less than for those under 65 (2.94 versus 3.20, on a scale of 1–7). In the same way, while people over 65 think it is more important to increase the size of the pension than those under 65, the difference between the two groups is only −0.34 (3.82 versus 4.16, on a scale of 1–7). The small size of these effects is consistent with the international literature, and suggests that socio-demographic characteristics are not the dominant determinants of preferences over retirement income policies.
A feature of the international literature is the evidence that people have little tendency to answer surveys in a particularly self-interested manner. The finding that survey results only show weak evidence of self-interested behaviour was made using US data by Ponza, Duncan, Corcoran and Groskind (1988) and Sears and Funk (1990), and confirmed by Evans and Kelley (2004) and Lynch and Myrskylä (2009) using Australian and European data. These papers show that a respondent’s attitudes, not his or her income or age, are the most important determinants of retirement income survey responses. Boeri et al. (2002) argue there is somewhat stronger evidence of self-interested behaviour in two Italian and German surveys, but even in these surveys age and income provide little explanatory power of individual preferences.

The results from our survey show more evidence of self-interested responses than the results from other surveys, even if the size of the effects is small. In particular, we find a small tendency for people in lower-income households to oppose compulsion, to favour means-testing, and to favour a lower entitlement age. We also find a small tendency for people over 65 to favour higher pensions and a higher age of eligibility, and to be more opposed to means-testing. It is not clear why these results are more pronounced than those in other surveys, but it may reflect a more accurate approach to estimating preference rankings. It should be noted, however, that a majority of all demographic subgroups would be prepared to raise taxes immediately if that would reduce taxes on future generations, a result opposed to the notion that people respond in self-interested ways.

### 3.5 Cluster analysis

To investigate the diversity of preferences further, we used Lloyd’s (1982) algorithm to sort the respondents into five endogenously determined preference clusters. Each person is allocated to the cluster that has the nearest mean preference vector; as members of the cluster will have greater affinity with each other’s views than with members of other clusters, the mean Spearman rank correlations for the clusters are much higher than the mean rank correlation estimated for the whole sample, and range from 0.39 to 0.61. Each cluster contains between 13 per cent and 27 per cent of the respondents. We chose five clusters, as the estimated partitions in this case were stable.9

The five clusters are shown in Table 6. They differ largely by the way their members rank the age of eligibility, universality/means-testing, and saving flexibility/compulsion criteria. With one exception, each of these three criteria are ranked 1 or 2 (most important) or 6 or 7 (least important) in each cluster. People in the two largest clusters (cluster 1 – ‘Status quo plus compulsion’ and cluster 2 ‘Raise age plus compulsion’) concur that universality is the most important and that saving

---

9 The results of the cluster analysis are similar if people are allocated to four clusters, but the estimated partitions for three or for six cluster groups were not stable.
flexibility is the least important criteria, but disagree as to whether increasing the age of eligibility from 65 to 67 is the second most important or the second least important criterion. People in cluster 3 (‘Means-tested redistribution’) favour the introduction of a means-test to fund higher pension payments, and also support compulsion. People in cluster 4 (‘Pension minimalists’) favour the least government intervention – they are against means-testing and compulsion, and want the age of eligibility increased. The fifth cluster (‘No compulsion’) is the smallest and least cohesive cluster and comprises people who are unified because they strongly favour saving flexibility rather than compulsion.

Table 6. The five cluster groups

<table>
<thead>
<tr>
<th>Central vector</th>
<th>Cluster 1 Status quo + compulsion</th>
<th>Cluster 2 Raise age + compulsion</th>
<th>Cluster 3 Means-tested redistribution</th>
<th>Cluster 4 Pension minimalists</th>
<th>Cluster 5 No compulsion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample fraction (s. dev)</td>
<td>27% (0.3%)</td>
<td>24% (0.9%)</td>
<td>18% (0.7%)</td>
<td>17% (0.7%)</td>
<td>14% (0.7%)</td>
</tr>
<tr>
<td>Mean Spearman correlation</td>
<td>0.56</td>
<td>0.61</td>
<td>0.48</td>
<td>0.44</td>
<td>0.39</td>
</tr>
<tr>
<td>Mean minimum distance</td>
<td>16.6</td>
<td>14.1</td>
<td>20.6</td>
<td>19.0</td>
<td>21.7</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations derived from the survey responses.

Can membership of a cluster be predicted from a person’s socio-demographic characteristics? In general, no. For most population subgroups it is not possible to reject the hypothesis that the subgroup has the same distribution across the clusters as the overall sample. There are two main exceptions. First, non-Europeans, particularly Māori, are more likely to be in cluster 5 (‘No compulsion’) and less likely to be in cluster 2 (‘Raise age plus compulsion’) than Europeans; 23 per cent of Māori were in cluster 5 relative to the national average of 14 per cent. Secondly, respondents in high-income households are 4 percentage points less likely to be in cluster 3 (‘Means-tested redistribution’) and cluster 5 (‘No compulsion’) than the general public.

In contrast, there are significant differences in the allocation across clusters of the groups that self-identify in terms of their level of confidence that they will have a comfortable retirement. People who do not expect to be comfortable are significantly more likely to be in clusters 3 and 5 (‘Means-tested redistribution’ and ‘No compulsion’) and less likely to be in cluster 4 (‘Pension minimalists’). Those who are confident they will be comfortable in retirement are much more likely to be in cluster 4 (‘Pension minimalists’) and much less likely to be in cluster 3 (‘Means-tested redistribution’). These differences reinforce the finding that differences based on non-observable characteristics tend to be larger than differences based on observable characteristics.
4 Policy choices

A key advantage of the multi-criteria decision analysis technique associated with the 1000Minds software is that the estimates of individual preferences can be used to evaluate how individuals will rank different policy packages. This evaluation is possible because the preference relationships of each respondent are used to estimate a cardinal utility function. The policy evaluations are made by (i) categorising each policy according to the survey criteria; (ii) using the respondent’s estimated preference weights to estimate his or her utility corresponding to the outcomes of each policy; and (iii) ranking each policy option by the utility scores.

In this section we estimate New Zealanders’ preferences over three fiscally neutral retirement income policies that have been extensively discussed in New Zealand. The first policy, ‘PAYGO65’ is the continuation of New Zealand Superannuation in its current form. It has the following features: (i) The pension is universal, and not means-tested. All people satisfying residency criteria receive the pension when they turn 65. (ii) The weekly payment for single recipients is approximately $360 in 2014–15, and adjusted for wage inflation. (iii) The pension is largely funded on a pay-as-you-go basis. No tax increases are required to finance current payments, but taxes will have to be increased to finance payments as the population ages. (According to Treasury projections, taxes will need to increase by 4.6 per cent of taxable income by 2050, a number we use as a proxy for the tax on future generations.) The second policy, ‘PAYGO67’ is similar but the age of eligibility for people born after 1953 is increased to 67 from 2020. No change in current taxes is necessary, but the taxes needed to pay for the pension will only increase by 3.5 per cent of taxable income by 2050. The third policy, ‘SAYGO65’, keeps the age of eligibility and the structure of pension payments the same as PAYGO65, but a tax surcharge equal to 2 per cent of taxable income is immediately imposed and placed in a sovereign wealth fund, the New Zealand Superannuation Fund. The additional contributions are assumed to cease after 30 years, at which point the ongoing earnings of the fund will reduce the increase in future taxes to 2.6 per cent of taxable income. The differences in the three policies are shown in Table 7.

---

10 The New Zealand Superannuation Fund was established in 2002 for this purpose.
11 We assume (i) 2 per cent of taxable income is added to the New Zealand Superannuation Fund each year; (ii) income increases at 2 per cent per year; and (iii) the Fund compounds at a 4 per cent real rate of return. After 30 years the accumulated Fund will have increased by 84 per cent of the contemporaneous level of income, and will produce ongoing returns that allow taxes to be reduced by 2 per cent of contemporaneous income. These assumptions satisfy the wording of the fourth criterion. If the rate of return is lower than 4 per cent real, the contribution period would have to be lengthened. As a 4 per cent rate of return is low by historic standards the contribution period could be shorter than 30 years.
Table 7. Three possible retirement income policies

<table>
<thead>
<tr>
<th>Policy</th>
<th>Age of eligibility</th>
<th>Tax increase in 2015</th>
<th>Tax increase in 2050</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAYGO65</td>
<td>65</td>
<td>0%</td>
<td>4.6%</td>
</tr>
<tr>
<td>PAYGO67</td>
<td>67</td>
<td>0%</td>
<td>3.5%</td>
</tr>
<tr>
<td>SAYGO65</td>
<td>65</td>
<td>2%</td>
<td>2.6%</td>
</tr>
</tbody>
</table>

Source: Authors’ summary of New Zealand retirement policies.

When the policy outcomes do not exactly coincide with the survey categories, the utility function is interpolated. The accuracy of the rankings increases with the number of categories in each criterion, but is highest when the policy outcomes exactly coincide with the survey categories, for then interpolation of the utility function is unnecessary. For this reason, the comparison of the SAYGO65 and the PAYGO65 policies is most accurate as the differences in the two policies (the differences in the current and future tax rates) exactly coincide with the differences in the survey tax criteria categories.

Table 8 shows the proportion of the sample that ranks each of the three policy options the highest, second highest, and lowest. The table shows that 58 per cent of respondents would obtain the greatest utility from the SAYGO65 policy, 26 per cent from the current PAYGO65 policy, but only 16 per cent from the PAYGO67 policy. Furthermore, only 16 per cent of the population would have the lowest utility from the SAYGO65 policy, whereas 28 per cent and 56 per cent would have the lowest utility from the PAYGO65 and PAYGO67 policies respectively. These results clearly indicate that, judged in terms of their own preferences, the policy option of raising the age of eligibility is the policy least preferred by the largest number of people and most preferred by the smallest number. In contrast, a policy of maintaining the age of eligibility at 65 and prefunding some future New Zealand Superannuation payments is the policy most preferred by the largest number of people and least preferred by the fewest. These results reflect the high rank most respondents place on the importance of avoiding large tax increases on future generations: 65 per cent of all respondents indicated they would accept an increase in current taxes equal to 2 per cent of taxable income if it would reduce taxes on the next generation by a similar amount. This preference ranking, which was shared by almost all population subgroups, is one of the strongest findings of the survey.

---

12 For example, the utility corresponding to a 4.6 per cent increase in future taxes is calculated by multiplying the weights associated with a 3 per cent and a 5 per cent increase by 0.2 and 0.8 respectively.
Table 8. Welfare ranking of different policies

<table>
<thead>
<tr>
<th></th>
<th>Highest ranking</th>
<th>Middle ranking</th>
<th>Lowest ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAYGO65</td>
<td>26%</td>
<td>46%</td>
<td>28%</td>
</tr>
<tr>
<td>PAYGO67</td>
<td>16%</td>
<td>28%</td>
<td>56%</td>
</tr>
<tr>
<td>SAYGO65</td>
<td>58%</td>
<td>26%</td>
<td>16%</td>
</tr>
</tbody>
</table>

Note: The table shows the fraction of the population giving each policy the 1st, 2nd or 3rd highest ranking.

Source: Authors’ calculations derived from the survey responses.

Does this ranking differ across population subgroups? Given the earlier finding that there are only small differences in the preferences of most subgroups, it is not surprising that different population subgroups rank the policies in a very similar manner. The SAYGO65 policy is the most popular and the least unpopular policy option for all the population subgroups analysed, and the PAYGO67 policy is the least popular and most unpopular option.13 Unsurprisingly, however, the popularity of the policies varies substantially across the preference clusters. The PAYGO67 policy option is relatively popular with members of cluster 4 (‘Pension minimalists’) but with almost no-one else, and it is overwhelmingly the least popular choice with members of clusters 1 (‘Status quo plus compulsion’), 3 (‘Means-tested redistribution’) and 5 (‘No compulsion’).

While we have only formally evaluated three policies, the advantage of this approach is that many others could be evaluated. For example, we could evaluate whether people would be willing to increase current and future tax rates to increase the size of the pension by $30 per week, or we could analyse the effect of indexing future amounts of New Zealand Superannuation to the consumer price index rather than wages. It is also possible to evaluate how policies that provide people with a range of options might affect welfare. An example of such a policy is the so-called flexible-Superannuation option, which is available in countries such as the United States that allows people to defer receiving a pension in exchange for receiving a larger amount. In a New Zealand context, we can evaluate the benefits of providing people with the option of receiving an extra $30 per week if they delay taking a pension until age 67, assuming that this was fiscally neutral.14 Our survey indicates that this option would be the preferred choice for 46 per cent of respondents, while the remaining 54 per cent would prefer current arrangements. This suggests that a policy that allowed people to choose when they got the pension would improve the welfare of 46 per cent of the population.

13 The Asian ethnicity group is the only population subgroup in which a majority of people did not rank the SAYGO65 policy as the most preferred, and even for this group it was still the single most preferred policy.
14 The policy could be evaluated if it were not fiscally neutral by calculating the changes in taxes needed to fund it.
5 Conclusion

This paper investigates whether a multi-criteria decision-making survey can be used to help develop core public finance policy options, in this case retirement income policy. The answer is ‘yes’. The survey provided detailed information about the relative importance of specific policy criteria. It shows, for example, that having a universal pension without a means-test is the most important issue for the greatest number (42 per cent) of respondents, while the least important issue for the largest number of respondents (38 per cent) is saving flexibility and compulsion. Respondents also revealed strong preferences about the timing of the taxes necessary to pay for pensions, with a large majority opposed to large tax increases on future generations. In contrast, there is considerable disagreement about the desirability of raising the age of eligibility from 65 to 67, with equal numbers of people either strongly opposed to or unconcerned about the policy. All of these findings can be disaggregated by population subgroups. We show, for example, that while few people overall would be opposed to the introduction of a small compulsory saving scheme, there is more opposition from people living in low-income households and from Māori.

The results indicate that while preferences about retirement income are diverse, they depend only a little on observable characteristics such as age, education, income or ethnicity. Rather, New Zealanders’ preferences reflect unobservable characteristics and people can be systematically grouped into clusters who share similar attitudes. Some of these attitudes appear to reflect their expectations about their level of comfort in retirement. However, it is likely that these attitudes also reflect deep-seated philosophical approaches to life (Bowles and Gintis, 2000). For example, there seem to be an identifiable cluster of people united in a preference for minimalist government intervention into retirement income policy, and another that is keen on greater redistribution.

An advantage of this survey technique is that the relative merit of complex policy proposals can be evaluated on the basis of individual rather than average preferences. We used the technique to make an assessment of three retirement income policies at the heart of contemporary debates in New Zealand: the current policy, a variation in which the age of eligibility is raised by two years to reduce future tax obligations, and another variation in which current taxes are increased to reduce future tax obligations. By estimating individual preferences, we show that raising the age of eligibility by two years is the most preferred policy for the fewest number of people and the least preferred policy for the largest number. In contrast, a policy that raises current taxes to prevent even larger future tax is preferred by a majority of the population and, indeed, a majority of almost all population subgroups.
Multi-criteria decision surveys, like other surveys, have some limitations. Surveys need to have enough criteria to capture the real-life complexity of an issue. Survey respondents must be able to understand the implications of different category levels of each criterion, and this may rule out many potential survey topics. Surveys are subject to framing biases, although to address framing concerns in future research it would be possible to use several differently framed surveys, possibly applied to smaller samples. Nonetheless, this survey provided an astonishing amount of information from a large number of people about the relative importance of several different aspects of retirement policies. Consequently, on the basis of the results obtained from this research, we believe the multi-criteria decision analysis approach has considerable potential to improve policy analysis and development.

References


ARGUMENT
Government failure and success: A trans-Tasman comparison of two insulation subsidy schemes

Nicholas Preval, Jenny Ombler, Arthur Grimes, Michael Keall and Philippa Howden-Chapman

In the wake of the 2008 Global Financial Crisis (GFC), the governments of Australia and New Zealand undertook a variety of economic stimulus measures, including home insulation and heating retrofit programs. Australia’s Home Insulation Program (HIP) ended early and in disarray (Hawke, 2010; Kortt and Dollery, 2012) while New Zealand’s Warm Up New Zealand: Heat Smart (WUNZ:HS) program was considered a success, outperforming agreed targets and time frames and producing a variety of health and other benefits (Grimes et al., 2011, 2016; O’Sullivan, Barnard, Viggers and Howden-Chapman, 2016; Preval, Keall, Telfar-Barnard, Grimes and Howard-Chapman, 2017).

Effectively, across these two related schemes, one resulted in government failure (Le Grand, 1991) while the other proceeded as well as (or better than) expected. We examine key proximate and more fundamental reasons behind this differential experience. The paper identifies differences in the key characteristics of the two programs and the political and policy contexts that could explain such differing outcomes. These include differences in governance, program targets, program design, implementation, evaluation and the influence of prior public health research.

While economic stimulus was an aim in each case, our analysis shows that such stimulus needs to be informed by prior research and/or prior implementation experience, especially if the program is to be implemented quickly. The lack of each of these factors in the Australian HIP case meant that its (rushed) decisions were poorly informed, resulting in failure of both the insulation and stimulus aims of

1 npreval@allenandclarke.co.nz, University of Otago and Allen and Clarke Policy and Regulatory Specialists; University of Otago; Victoria University of Wellington and Motu Research; University of Otago; University of Otago.
the scheme. By contrast, the New Zealand WUNZ:HS scheme was underpinned by significant prior research and implementation experience, meaning that the stimulus decision rested on an appropriate information base.

The paper first provides background to the two schemes, including the program structures and details about product choices and installation methods. These sections set out the proximate causes of failure versus success of the two schemes. We then analyse the underlying causes of the differing results across countries by analysing the processes of policy advice and consultation, and of implementation and evaluation of the schemes. Some concluding comments highlight our key findings.

**Background**

When the 2008 GFC resulted in the collapse of non-bank financial institutions and dampened demand in sectors such as construction, both Australia and New Zealand’s central governments responded with a variety of Keynesian economic initiatives intended to stimulate growth and reduce unemployment. Both developed or accelerated home energy efficiency retrofit programs. In addition to providing stimulus, these programs were intended to address the high proportion of homes that did not meet minimum insulation standards: an estimated 700,000–900,000 homes in New Zealand were under-insulated (Johnson, Howden-Chapman and Eaqub, 2018), and 2.2 million homes lacked insulation entirely in Australia (Hanger, 2014).

New Zealand’s WUNZ:HS was a four-year (2009–2013), NZ$347-million program administered by the Energy Efficiency and Conservation Authority (EECA) that was designed to partially fund a range of home insulation and heating retrofits; primarily ceiling and floor insulation and improved heating. The National government program was a development of an earlier, ongoing scheme called ‘Energywise Homes’ that had been instigated by the previous Labour Government and had been operational since 2003. As a response to the GFC, the stimulus potential of the program led to its expansion into WUNZ:HS with greater investment from the government. Ministerial and Cabinet papers leading to the decision reiterated energy and health co-benefits that underpinned the Energywise program, and emphasised the ways in which greater investment would also provide a stimulus effect (Hull, 2009; McNicholas, 2009). Based on prior research, WUNZ:HS was predicted to reduce energy expenditure and also to produce public health co-benefits resulting from increases in indoor temperatures and reductions in mould and allergens in houses receiving insulation and/or heating, which had been established by randomised community trials in New Zealand (Chapman, Howden-Chapman, Viggers, O’Dea)

---

2 These figures are estimates, as neither country conducts regular random, nation-level housing quality surveys.
and Kennedy, 2009; Howden-Chapman et al., 2007, 2008). Subsequent evaluations confirmed that the program produced both energy savings (Grimes et al., 2016) and health benefits (Preval, 2012, 2015; Preval et al., 2017; Telfar Barnard et al., 2011). The program was subsequently continued as Warm Up New Zealand: Healthy Homes with reduced funding (EECA, 2013) and has more recently been continued under the incoming Labour Government.

Australia’s HIP was a two-and-a-half-year, AU$2.7-billion federal program administered by the then Department of the Environment, Water, Heritage and the Arts (DEWHA). The scheme part-funded home ceiling insulation retrofits. While the scheme was predicted to have some energy efficiency co-benefits, it was primarily focused on economic stimulus (Auditor General, 2010). The program was derived from a set of earlier proposals to improve energy efficiency, as part of a wider Carbon Pollution Reduction Scheme. Where these proposals discussed insulation, they emphasised job creation, with energy saving co-benefits. They had not been developed into policy prior to the development of the HIP (Hanger, 2014, p. 6). The HIP was trialled in early 2009 (Phase One) before being rolled out in July 2009 (Phase Two). During its brief period of operation, four workers lost their lives while undertaking the work and the program was beset with claims of fraud and poor-quality work which ultimately required costly remediation. The HIP was formally ended in February 2010 after months of controversy. Extensive investigation of the HIP, which included four government-commissioned reports (Auditor General, 2010, Hanger, 2014; Hawke, 2010; Senate Committee, 2010) and three peer-reviewed articles (Dollery and Hovey, 2010; Kortt and Dollery, 2012; Lewis, 2010), identified numerous design and implementation failures.

Table 1 summarises the key characteristics and timelines of the two programs.

Table 1. Key program characteristics

<table>
<thead>
<tr>
<th></th>
<th>HIP</th>
<th>WUNZ:HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>completion date</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phase Two (1 Jul 2009 – 19 Feb 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phase Three (19 Feb 2010 – remediation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>only)</td>
<td></td>
</tr>
<tr>
<td>Program funding</td>
<td>AU$ 2.7 billion (reduced to</td>
<td>NZ$347 million</td>
</tr>
<tr>
<td></td>
<td>AU$2.45 billion, 2 Nov 2009)</td>
<td></td>
</tr>
<tr>
<td>Primary program</td>
<td>DEWHA</td>
<td>EECA</td>
</tr>
<tr>
<td>implementation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## HIP WUNZ:HS

<table>
<thead>
<tr>
<th>Intended outcome (actual outcome)</th>
<th>Improved ceiling insulation for 2.2 million homes (approximately 1.16 million completed)</th>
<th>Retrofitted insulation and space heating for 188,000 homes (approximately 234,000 completed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted co-benefits</td>
<td>Economic stimulus:&lt;br&gt;Reductions in energy use and energy efficiency improvements&lt;br&gt;Reduced CO2 emissions&lt;br&gt;Improved comfort</td>
<td>Health savings:&lt;br&gt;Economic stimulus&lt;br&gt;Reductions in energy use and energy efficiency improvements&lt;br&gt;Reduced CO2 emissions&lt;br&gt;Improved comfort</td>
</tr>
<tr>
<td>Retrofits available under scheme</td>
<td>Ceiling insulation</td>
<td>Ceiling and floor insulation (mandatory to have both if possible), vapour barriers, draught excluders, hot water pipe lagging, cylinder wraps; Clean heating (either a heat pump, wood burner, pellet burner or a reticulated gas heater)</td>
</tr>
<tr>
<td>Funding available to participants/consumers</td>
<td>AU$1,600 per home for ceiling insulation (reduced to AU$1,200, 2 Nov 2009), paid by central government</td>
<td>33% central government funding for ceiling and floor insulation, draught proofing, moisture barriers and remediation (maximum NZ$1,300), (60%* funding for low-income households or landlords of low-income households – no maximum) NZ$500 central government funding towards clean heater ($1200* if low-income household); funding for clean-heaters eventually phased out of program * Additional funding/more complex funding arrangements available from charities, local government or lines companies</td>
</tr>
<tr>
<td>Service delivery model</td>
<td>Participants/consumers contracted directly with installer.&lt;br&gt;Phase 1 (3 Feb 2009 – 30 Jun 2009): Claim required two quotes and a site inspection (Senate Committee, 2010). Money was paid upfront by participants/consumers who then claimed a rebate processed and paid by DEWHA.&lt;br&gt;Phase 2 (1 Jul 2010 – 19 Feb 2010): Installers were paid rebate amount directly following claim via Medicare Australia. Consumers often did not pay any money at all.</td>
<td>Contract between consumers and service providers.&lt;br&gt;Consumers paid service providers directly (less rebate), service providers obtained rebate from EECA via claim. Consumers were able to pay via mortgage extension, interest-free loan, or rates. Consumers in some cases had access to additional funding from third parties such as local government and charity organisations.</td>
</tr>
</tbody>
</table>

Note: CO₂=carbon dioxide.

Program structure

Many of the proximate failures of the Australian HIP can be linked directly to program design choices such as minimal program entry requirements for installers, and a rebate structure which did not require home owner contributions. It was relatively easy for installer organisations to register under the HIP as they were only required to have a valid Australian Business Number and to agree to the Terms and Conditions of Registration; 7,841 completed at least one installation (Senate Committee, 2010). The quick uptake limited the ability of Medicare to confirm that registered installers met program entry requirements in a timely manner, and meant that a number of dishonest operators were able to enter the HIP and operate unchecked, either billing for work not completed, or otherwise deliberately circumventing program regulations. Hawke (2010, p. viii) cites, ‘disturbing claims about the high level of fraud perpetrated by unscrupulous operators. Despite some safeguards against fraud, no one foresaw the possible extent of potential malfeasance which was simply alarming’.

An equal concern was the retrofitting of insulation by unqualified and/or inexperienced installers, which was linked to poor-quality workmanship, dangerously installed insulation and to the deaths of four installers, which ultimately led to the end of the HIP. Installers needed only generic health and safety training and to be supervised by somebody with: ‘prior industry experience; a trade qualification in an approved trade; or insulation specific training’ (Auditor General, 2010, p. 104). The decision to require only supervisors to meet minimum competency standards was made by DEWHA in consultation with a range of stakeholders, and the trade-off between requiring minimum standards for all workers and just supervisors was explicitly considered; however, it was decided that ‘it would not be possible to implement, given the short period of time available and the need for low barriers to entry to the program’ (Auditor General, 2010, p. 107). It was later reported that supervision was often minimal, meaning that unqualified and inexperienced installers were placed in very difficult situations (Lewis, 2010).

By contrast, there were only about 60 service providers under the WUNZ:HS program (EECA official, personal communication, 2010). The average WUNZ:HS service provider retrofitted approximately 780 homes per year over the course of the program, which contrasts with approximately 150 per year for installers under the HIP. WUNZ:HS service providers were required to complete an annual review as a condition for contract extension, and acceptance of stringent auditing requirements (EECA, 2009). Small operators, such as those who thrived under the HIP, were unlikely to have gained entry to the WUNZ:HS program, given the extensive tendering process, while any service provider that did gain entry to the program would have more to lose if found to be engaging in fraudulent practices. WUNZ:HS did not explicitly set minimum installer competencies, but the
request for proposals stated that when considering proposals 35 per cent weight would be given to ‘capability’, which included ‘proven personnel and experience or capability in installing EECA approved products’ and ‘evidence of ability to complete work to the Standards outlined in EECA’s Audit and Quality Manual’ (EECA, 2009, p. 12). Even with these barriers to entry, audits during the early months of WUNZ:HS’s operation found a problematic failure rate for a mixture of administrative and safety reasons. It was, however, possible for EECA to swiftly address these issues with higher rates of auditing and appropriate communication with the 60 service providers.

The other fundamental difference in the design of the two programs was the fact that Phase Two of the HIP did not require a household contribution unless the total cost of an insulation retrofit exceeded a preset limit, with rebates paid directly by the government. Because the majority of HIP retrofits were below the threshold for household contribution, most households had limited motivation to make informed decisions about which installation company to choose, and to monitor the quality of work (which is difficult even when householders are motivated due to information asymmetries in this industry) (Senate Committee, 2010). As with many of the other problematic aspects of the Australian HIP program design, there was awareness of these issues among senior officials during the design phase, including explicit mention of these issues in the Minter Ellison risk register published in April 2009 (Senate Committee, 2010). By contrast, the WUNZ:HS rebate structure, which required home owners or third parties to contribute financially, can be assumed to have encouraged a higher degree of consumer vigilance.

Product inclusion and installation methods

Although the two programs had different scopes, the common element of both was insulation. The treatment of foil insulation is particularly interesting as a basis of comparison. Foil insulation is an insulation material that reflects a high proportion of radiant heat. Foil insulation was available under the HIP. Tragically, electrocution killed three installers who inadvertently pierced live wiring with metal staples during foil insulation installation (Senate Committee, 2010), and poorly installed foil retrofits meant that many ceilings were potential fire or electrocution hazards, later requiring checking, modification or removal of the insulation retrofitted (Auditor General, 2010). Industry consultation prior to the start of the program had raised awareness of the risks of foil insulation but, as was also the case with proposed minimum installer qualifications, some members of the established insulation industry opposed changes to current practice that included the use of metal staples, citing the lack of previous issues with these products (Hawke, 2010). The Senate Committee report concluded that it was the combination of inexperienced installers and products requiring skilled installation that led to the foil-related safety issues.
Foil insulation was not available under the WUNZ:HS, although foil floor insulation had been available under a previous EECA program until mid-2008. In 2007, several New Zealand installers died of electrocution while installing foil insulation (not under EECA administrated programs), and both safety and other issues such as a lack of durability led to the eventual exclusion of foil from EECA-funded programs including WUNZ:HS. During a stakeholder meeting early in Phase One of the HIP, the New Zealand deaths were discussed, but this and other evidence did not dissuade Australian officials from including foil in their program (Senate Committee, 2010).

Policy advice and consultation

The processes by which each of the HIP and WUNZ:HS policies were designed reflects different prior experience with insulation schemes, different policy environments and different research profiles. The HIP was designed prior to regulatory reforms in Australia, meaning that government programs did not then require a regulation impact statement or cost–benefit analysis (R. Reilly, Director of Office of Best Practice Regulation, personal communication, 12 March 2019). Since reforms in 2013, government programs are normally required to go through these processes, which may have identified some risks or limitations of the proposed program. Instead, policy decisions were made through the Strategic Priorities and Budget Committee (SPBC), a subcommittee of four members of Cabinet. The full Cabinet was briefed on the overall policy on the same day that the HIP was announced (Hanger, 2014, pp. 78–81). The SPBC was criticised as sidelining Cabinet while concentrating decision-making too heavily within the SPBC (Kefford, 2013).

The 2014 Royal Commission of Enquiry report on the HIP notes that there was a significant lack of relevant expertise among all who were charged with the design and implementation of the policy. Risk was identified early in the HIP policy design process. Subsequently, Minter Ellison Consultants were appointed as ‘external risk consultants’. The report suggests that accounting for risk failed in two primary ways. First, DEWHA had earlier decided that responsibility for compliance with health and safety legislation lay with the employer, not with the government. Second, the process undertaken by Minter Ellison Consultants with DEWHA was ‘deficient’ (Hanger, 2014, p. 119), and did not ensure that particular expertise on the risks surrounding this type of program was included. As a result of both of these failures, injury to installer was not included on the final risk register and was only reinstated after a death had occurred. Summarising issues around risk more generally, the report notes that frank advice about risk to senior officials and

---

3 Foil insulation was made illegal in 2016 as part of a review of the Residential Tenancies Act (NZ).
Ministers should include practical implications and risks for those directly involved in programs, policies or projects rather than a primary focus on reputational and political risks (Hanger, 2014, p. 310).

In assessing some of the decisions that undermined the HIP, the Senate Committee suggested that the Australian government ‘should establish a dedicated and industry-independent program to research insulation systems and help develop efficient and effective insulation policy’ (Senate Committee, 2010). However, while it is tempting to agree with the Senate Committee (2010) that Australia would have benefited from an ‘independent scientific facility in Australia able to research the properties of the various systems and advise on insulation policy in context of overall energy efficient housing goals’, the unwillingness of the federal government and senior officials to consider modifications that might have compromised the economic stimulus effect of the HIP led them to give little weight to the results of the consultation and risk assessment processes. This resulted in few changes in practice that might have mitigated risk, or have had a greater impact on co-benefits, largely due to the perceived delay this would have caused to the implementation of the program which was a primary concern (Hanger, 2014, pp. 25–26). The 2014 Royal Commission of Enquiry into the HIP noted: ‘careful planning was sacrificed to the perceived need for speed’ (Hanger, 2014, p. 28). Furthermore, the Enquiry highlighted the need for measures that empower public servants to disagree with politicians, a difficult task when the security of tenure of public officials has been eroded in Australia, leading to fear that insecure positions might be compromised if advice given is unpalatable to Ministers (Hanger, 2014). The Enquiry recommended cultural changes that would encourage public servants, and also educate Ministers on their role in receiving advice (Hanger, 2014, p. 307). Understanding why the stimulus goal overrode other considerations, which arguably contributed to program failure, leads us to look at the wider policy context of the two countries.

Insulation has widely been accepted as a means of improving population health in New Zealand and insulation programs have a long history. Two organisations fill the policy advice role identified by the Senate Committee in the New Zealand context. First, New Zealand has the Health Research Council–funded He Kainga Oranga/Housing and Health Research Programme, a publicly funded university-based research organisation, which has provided considerable independent, policy-relevant research on the value of insulation and heating retrofits. Second, New Zealand has the Building Research Association of New Zealand Ltd (BRANZ), an independent organisation funded by a levy on consented building projects which must be used to ‘promote and conduct building-related science and research … [including] testing and assurance for insulation products, and technical knowledge about how to correctly specify and install insulation and the consequences of poor installation’ (BRANZ, 2010, p. 15).
In particular, He Kainga Oranga’s research has had a great deal of influence in raising the perception of links between health and housing in New Zealand over the period 1996–2010 (Bierre and Howden-Chapman, 2017; Carroll, Blewden and Witten, 2008; Fougere, Howden-Chapman and Bennett, 2012). This rise coincides with publication of key results from randomised controlled trials demonstrating the health and energy co-benefits of retrofitted insulation (Chapman et al., 2009; Howden-Chapman et al., 2007). It is now widely recognised in New Zealand that there is a research-proven link between poor housing quality and health (Carroll, Blewden and Witten, 2008).

At the time that WUNZ:HS was being developed, the consensus view linking insulation and health was well established. Furthermore, prior to WUNZ:HS, there was considerable experience with implementing publicly subsidised insulation schemes with EECA having already implemented numerous insulation and energy efficiency retrofit schemes. Cross-party support for insulation retrofit programs was evidenced by the previous government’s planned NZ$1-billion insulation retrofit program, which was ultimately superseded by WUNZ:HS. New Zealand had also previously faced a leaky building crisis (Building Industry Authority, 2002), demonstrating the long-term consequences of poorly informed short-term decisions with long-term impacts on housing quality and health (Douwes and Howden-Chapman, 2011; Howden-Chapman, Ruthe and Crichton, 2011).

By contrast, we found little research about home insulation and its benefits in the Australian context. Savings were identified as co-benefits resulting from the HIP but there was little discussion of any health co-benefits (Chapman, Preval and Howden-Chapman, 2017; Howden-Chapman and Chapman, 2012). While Australia is a drier and warmer country than New Zealand on average, analysis by Cheng et al. (2010) indicates that its citizens can experience hazardously high and low indoor temperatures (Basu, 2009; Pierse et al., 2013), both of which can be ameliorated by insulation. Given the value placed on the co-benefits of insulation, which have been internationally accepted (International Energy Agency, 2015; World Health Organization, 2011), and the large body of institutional knowledge and previous uptake within New Zealand, greater weight appears to have been placed on quality rather than quantity of delivery in New Zealand relative to Australia.

**Implementation**

The HIP and WUNZ:HS were funded and implemented by federal and national governments respectively. Reviews of the two programs suggest that federal government administration was not ideal in the case of the HIP, as operational policies were typically carried out at state level. Kortt and Dollery (2012) note that the HIP was predicted to have both state and national benefits. Consistent with the
literature on subsidiarity (Wills, 2016), they postulate that central government has an advantage in terms of revenue-gathering while state government has a comparative advantage in the administering of programs that require local knowledge and local experience. This suggests that the HIP would have functioned better as a centrally funded, but state-implemented program, particularly given that many states had previous experience administering similar programs (Hawke, 2010; Kortt and Dollery, 2012). This argument is consistent with the many complaints and concerns that state governments voiced during both the design and implementation stages of the HIP (Lewis, 2010; Senate Committee, 2010).

By contrast, New Zealand has a unitary (as opposed to federal) system of government and at 20 per cent of Australia’s population, is closer in population size to that of a single Australian state. EECA, the implementing agency for WUNZ:HS, had considerable previous national experience administering similar programs and thus, using the criteria presented in Kortt and Dollery (2012), New Zealand’s centralised approach to both revenue-gathering and program administration is likely to have been preferable.

Although the degree of centralisation may have influenced the relative success of the two programs, many of the difficulties faced during the implementation of Australia’s HIP resulted from choices made during the design phase of the program. A key example is the massive increases in monthly retrofits during the period July 2009 – November 2009, which resulted from decisions designed to maximise the stimulus effect of the program. Given the high levels of demand, reports of poor workmanship and fraud could not be addressed in a timely manner as backlogs of complaints made the HIP’s monitoring and auditing program relatively ineffective (Auditor General, 2010).

In New Zealand, EECA also experienced difficulties in the early stages of WUNZ:HS, including capacity issues, but these difficulties did not overwhelm its systems. For example, EECA was able to improve poor-quality workmanship by raising external audit rates from 5 per cent to 10 per cent and in some cases demanding that offenders check all previous work. Such an approach would have been much harder to implement with many small providers and a greater rate of uptake.

While the two programs did have different rates of uptake, differences in program implementation were also influenced by differences in organisational experience. EECA’s long history of administering similar programs to WUNZ:HS gave it an advantage relative to DEWHA, which, as a policy development–based agency, had not implemented a program of this kind. The Senate Committee report noted that DEWHA did not have staff with experience in the insulation industry and that the management structure was not suitable for a program of the size of the HIP.
In Australia, in addition to capacity and experience issues, there was a clear reluctance by politicians and senior officials to take effective action to address problems during the implementation phase if this might compromise the speed of the rollout, again reflecting the overriding stimulus goal (Lewis, 2010). The Royal Commission Report notes that certain considerations were put aside, or inadequately attended to, in the interest of expediency. Two examples of this are the failure to exclude foil insulation, and a truncated training program, both of which were cited as decisions made in favour of the stimulus objective at the expense of health and safety. Furthermore, the report noted that the program’s expediency objective compromised ‘usual protections’ such as an audit and compliance regime (Hanger, 2014).

An argument could be made that stimulus in the face of economic slowdown is worthwhile in itself and in that respect the HIP scheme may be viewed as a success. However, stimulus can be achieved in multiple ways. A scheme characterised by poor workmanship led to both poor outcomes for home owners and discredited future insulation schemes, resulting in harms that would not have occurred with other forms of stimulus. The opportunity cost of using resources for the HIP rather than some other stimulus program is a real resource cost for the Australian economy.

Policy evaluation

A formal evaluation process was not built into Phase Two of the Australian HIP prior to the start of the program, although Phase One of the program did serve as a design and process evaluation tool, as did the various consultation processes. Ultimately, the swiftly commissioned independent Hawke Review served as a design and process evaluation for Phase Two, helping to finalise the suspension of the program; however, as noted in the Senate Committee report, ‘[t]he government’s move to commission an independent review of the HIP (the Hawke Review) was too little, too late and should have been undertaken earlier so that the findings could be used to improve the HIP’ (Senate Committee, 2010, pp. xiv). By contrast, the WUNZ:HS program included funding for an extensive outcome evaluation, which was timed so that results could lead to program modification or even end the program (Grimes et al., 2011). In the Australian case, it seems likely that the absence of a formal outcome evaluation process reflected the haste with which the HIP and other stimulus programs were rolled out.
Conclusion

There is a great deal that can be learned from this comparative case study of causes of government success and failure. The comparison of policy contexts illustrates the valuable contribution that prior experience with similar programs and prior national and international research can make in influencing key policy decisions. It is possible that research undertaken outside of the country carried insufficient weight to influence Australian policymaking. However, the failure of the HIP demonstrates the limits of research, especially given the dearth of prior policy-related experience, in the face of the hasty pursuit of a politically driven goal. The focus on expediency undermined the quality of the program. Ultimately, the cancellation of the scheme truncated ongoing stimulus effects.

A government focus on quantity rather than quality of provision, coupled with a poor information base and poor governance arrangements (especially federal government involvement in what is more properly a state government initiative), can – and did – result in a failed program. The comparison with the contemporaneous successful program in New Zealand shows that such government failure is not inevitable for this type of scheme. A key lesson is that a research-informed scheme that is informed by prior experience and that builds in appropriate evaluation steps is likely to outperform a simple stimulus-driven package. The result is a successful and sustainable stimulus package that achieves long-term health benefits for citizens.

References


