Measuring and Decomposing the Income Mobility of Individuals in New Zealand: Evidence from Administrative Data*

Nazila Alinaghi, John Creedy and Norman Gemmell†

Abstract

This paper uses administrative, longitudinal data on the New Zealand taxpayer population to examine the nature and extent of income mobility by individuals. It uses recently developed illustrative devices for mobility measures based on individuals' relative income growth and the extent of re-ranking within the income distribution over time, for periods of 1 to 15 years, during 2002 to 2017. Results highlight two phenomena not widely recognised hitherto. First, relative income growth is consistently higher (lower) for those with initially lower (higher) incomes, reflecting strong 'regression to the mean' processes. Second, there is a high degree of re-ranking by individuals within the income distribution over time: after 15 years, re-ranking of individuals' incomes represent around 30 to 45 per cent of the maximum re-ranking possible.

JEL Classification: D31, I32

Keywords: Income mobility

*This paper is part of a project on 'Measuring Income Inequality, Poverty, and Mobility in New Zealand', funded by an Endeavour Research Grant from the Ministry of Business, Innovation and Employment (MBIE) and awarded to the Chair in Public Finance at Victoria University of Wellington.

†Victoria University of Wellington

Disclaimer

Results reported below are based in part on tax data supplied by Inland Revenue to Statistics New Zealand (SNZ) under the Tax Administration Act 1994 for statistical purposes. Any discussion of data limitations or weakness is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements. Access to the data used in this study was provided by SNZ under conditions designed to give effect for the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the authors, not SNZ or individual data suppliers. These results are not official statistics. They have been created for research purpose from the Integrated Data Infrastructure and/or Longitudinal Business Database which are carefully managed by SNZ. More information about these databases can be obtained at: *https://www.stats.govt.nz/integrated-data/*.

1 Introduction

This paper uses administrative, longitudinal data on the taxpayer population to examine the nature and extent of income mobility by individuals in New Zealand over the period 2002 to 2017. The construction of the special dataset has been made possible due to the improved availability of anonymised administrative register data, such as from individuals' tax records, in New Zealand's Integrated Data Infrastructure (IDI). These administrative data sources provide several advantages compared with sample surveys. Administrative data have very large sample sizes, improved coverage of top incomes, avoidance of survey respondent dropout or attrition, and less measurement errors. While recognising the limitations of such data, for example the absence of information on non-taxable income, the dataset used in this paper nevertheless provides the most comprehensive information to date on NZ taxpayers' incomes, suitable for inequality and mobility analysis.[1]

The focus of the paper is on the construction of diagrammatic devices which succinctly convey the nature of what is clearly a highly complex dynamic process of income mobility among many thousands of individuals. Emphasis is given to two different aspects of income mobility. The first is concerned with relative income growth, while the second examines positional, or re-ranking, changes taking place within cohorts over time. In describing the nature of mobility, no attempt is made here to distinguish changes which are regarded by the individuals concerned – or indeed policy-makers – either as desirable or undesirable.[2]

The value of diagrams to summarise income distribution characteristics is of course exemplified by the famous Lorenz curve, which has become a standard device to illustrate the nature of cross-sectional income distributions. With individual observations

[1]Income mobility is of course only one aspect of more general social mobility, including intergenerational, as well as intra-generational, mobility. While income mobility is relatively easy to measure and quantify, it does not seek to capture broader dimensions of mobility such as that associated with changes in social class, educational or occupational status; see Simandan (2018), the contributions by Atkinson and Goldthorpe in Svallfors (2005) and Markandya (1982) for further discussion.

[2]Following Fields (2000), a number of authors have pointed to the normative ambiguity associated with (possibly desirable) flexibility in long-term income movements versus (undesirable) short-term volatility. Jäntti and Jenkins (2015) suggest that the concept of income risk can be regarded as one component of longer term income inequality. In this view, changes in an income inequality measure over time have both permanent predictable, and transitory unpredictable, components. They label the latter as 'income risk'.

arranged in ascending order, the Lorenz curve plots (within a box of unit height and base) the cumulative proportion of total income against the corresponding cumulative proportion of individuals. This provides much more information 'at a glance', about relative income inequality, than either the density function or the distribution function alone, and can quickly allow qualitative comparisons between different periods or population groups.

A challenge arises in the context of income mobility, where the same individuals are observed in, say, two different years and where the 'basic data' are in the form of a joint distribution. Three-dimensional graphs would be needed to plot such distributions and would not easily reveal the nature of relative income changes. One tabular approach to summarising the characteristics of such a joint distribution involves the construction of transition matrices for movements between, say, deciles of the distributions, thereby compression a vast amount of information into a ten-by-ten table.[3] Such matrices are explored using the New Zealand individual data in Alinaghi *et al.* (2022b).[4]

A different approach involves specifying a simple dynamic process using a regression model with a small number of easily-interpreted parameters. This necessarily requires strong assumptions about the structure of income changes, particularly the form of conditional income distributions for the second period, for given incomes in the first period. The results of modelling of this kind of model for NZ individual incomes are reported in Creedy *et al.* (2021). Such a parsimonious specification is particularly useful when it is required to include income dynamics in wider economic models.[5]

Where, as in the present context, it is required to illustrate the main characteristics of mobility in a simple diagram, a number of alternatives have been proposed. Most of these have focused on income growth measures, conditional on initial incomes. These are described briefly in Section 2. However, the empirical analyses reported here make

[3]The use of deciles is of course only one of many choices to be made regarding income classes. For example, some studies use classes of equal absolute income ranges, while others use class widths with equal logarithmic ranges. Jäntti and Jenkins (2015) discuss various positional change mobility indices based on deviations from the diagonal of the transition matrix.

[4]Motivated by a suggested lack of transparency of transition matrices, Trede (1998) proposed a diagram showing profiles of various quantiles of conditional distributions of income in the second period, given income in the first period. For further discussion, see Creedy and Gemmell (2017).

[5]If concern is focused on the extent to which a summary measure of inequality varies as the accounting period is lengthened, then a simple two-dimensional diagram can be used, as illustrated for New Zealand in Creedy *et al.* (2021). The variation in such a measure has, following Shorrocks (1978), often been used as a summary measure of mobility.

use of several new types of diagram introduced by Creedy and Gemmell (2019a). First, with individuals ranked in ascending order of initial income, they defined a modified growth curve which plots the cumulative proportional income change *per capita* (not, as in previous growth curves, per head of the cumulated sub-group), against the corresponding proportion of individuals. This diagram enables three characteristics of mobility – incidence, intensity and inequality – to be clearly illustrated: it is referred to as a 'Three Is of Mobility', or TIM, curve, following the terminology adopted by Jenkins and Lambert (1997) in the context of cross-sectional poverty. Second, and in the context of positional mobility, Creedy and Gemmell (2019a) introduced a 'cumulative re-ranking curve', which considers the cumulative observed re-ranking across individuals, ranked in ascending order of their position in the initial income distribution. Third, they defined a 're-ranking ratio' (RRR) curve, which compares the ratio of observed re-ranking to the maximum feasible re-ranking for each individual (since the maximum differs across individuals).[6]

The TIM curve concept is described briefly in Section 3, and the positional change mobility diagrams are defined in Section 4. These new devices are applied to the special longitudinal dataset of individual taxpayers in New Zealand, summarised in Section 5. Section 6 reports results for TIM curves for taxpayers as a whole and for various subgroups. This is followed by results for re-ranking measures in Section 7. Conclusions are in Section 8.

[6] The various illustrative devices avoid an attempt to produce an overall measure of mobility. A simple approach, for example, would involve the proportion of off-diagonal entries in a transition matrix. Shorrocks (1978) proposed a mobility measure in terms of 'the degree to which equalisation occurs as the observation period is extended' (p.386). Using New Zealand taxpayer income data from 1994 to 2012, Creedy and Gemmell (2019a) and Creedy *et al.* (2021) report reductions in Gini and Atkinson inequality indices as the accounting period is lengthened from one year to up to 19 years (from different starting dates). Alinaghi *et al.* (2022a) perform a similar exercise using more recent and more comprehensive taxpayer data.

2 Income Growth Curves

This section briefly reviews alternative income growth curves used to illustrate mobility, clarifying the distinction between these approaches and the TIM curves used here.

2.1 Growth Incidence Curves

The 'growth incidence curve' (GIC) plots the income growth rate between two periods of each quantile or percentile of the distribution of initial incomes. As originally proposed by Ravallion and Chen (2003), the GIC is based on cross-sectional distributions for two periods and is therefore not capable of illustrating individual-specific income mobility.[7] Bourguignon (2011) extended the concept to capture longitudinal aspects of individual income growth in what he refers to as a 'non-anonymous growth incidence curve' (na-GIC). The absence of anonymity means that the same individuals are identified in both initial and 'terminal' income distributions.[8] The na-GICs are based only on the characteristics of the two relevant (longitudinal) distributions, but can easily display *relative* growth differences by subtracting overall income growth.

Beginning from an income distribution with a cumulative density function given by $F(y)$, the na-GIC is defined over both initial and terminal period distributions by first defining a 'quantile function', $y_F(p)$, as the inverse of $F(y)$. A similar function $y_\Phi(p)$, describes the equivalent terminal quantile function, conditional on initial incomes. Thus, income growth rates for each p^{th} percentile are given by:

$$g_{\Phi F}(p) = \frac{y_\Phi(p)}{y_F(p)} - 1 \tag{1}$$

and the 'distributional impact of growth is thus represented through the inverse of the cumulative density functions rather than those functions themselves' (Bourguignon, 2011, p. 609). A cumulative version of the na-GIC, referred to as the 'p-cumulative GIC', given by:

$$G_{\Phi F}(p) = \frac{\int_0^p g_\Phi(q) y_F(q) dq}{\int_0^p y_F(q) dq} \tag{2}$$

[7] A similar 'poverty growth curve' (PGC) to the Ravalion and Chen (2003) GIC was proposed by Son (2004), who illustrates *cumulative* growth across p percentiles of the income distribution. However, like Ravalion and Chen (2003), the PGC is based on comparisons of cross-sectional, rather than longitudinal, income distributions. See also Son (2007).

[8] Of course, in actual datasets, the observations for individuals are 'anonymised' and some kind of numerical identifier is used.

Graphical representations therefore involve plotting $g_{\Phi F}(p)$ or $G_{\Phi F}(p)$ against p.

2.2 Income Growth Profiles

Jenkins and Van Kerm (2016) define 'income growth profiles', IGPs, which are similar to those developed by Van Kerm (2009) and Bourguignon (2011).[9] They were largely concerned with the welfare dominance properties of individual income growth, based on an adaptation of the Atkinson and Bourguignon (1982) social welfare function where individual utilities are based on incomes in both the initial and terminal periods; see Jenkins and Van Kerm (2016, pp. 681-3).[10] Their objective therefore differs from the 'positive' description of income mobility properties pursued here. Nevertheless, their profiles capture two properties that are similar to the curves discussed in Section 3. The IGP plots a measure of average income growth, $m(p)$, for the pth percentile (or initial 'fractional rank'), against p, where in their case $m(p)$ is an expectation-based measure conditional on initial income. The IGP clearly bears a close resemblance to the na-GIC but does not require a common marginal initial income distribution.

Jenkins and Van Kerm (2016) also propose a cumulative version of the IGP (a CIGP) in which a measure of average income growth for those with initial incomes below $x(p)$ is plotted against p. The CIGP is given by:

$$C(p) = \frac{1}{p} \int_0^p m(q)dq \tag{3}$$

Thus, the CIGP plots areas below the income growth profile.[11]

These IGPs, CIGPs and the Bourguignon equivalents can be used to identify the *incidence* of mobility; that is, the mobility of a selected proportion, p, of the initial income distribution. They can also illustrate the *intensity* of mobility to some extent, for example by comparing the height of each CIGP at alternative values of p. However, some normalisation of CIGPs across periods would also be required for inter-period comparisons, where *relative* income growth is the relevant mobility concept. Identifying

[9] See also Grimm (2007).

[10] Palmisano and Peragine (2015) propose a similar welfare framework for analysing growth incidence. They argue that, unlike Bourguignon (2011) and Jenkins and Van Kerm (2011), their framework can incorporate horizontal inequality concerns.

[11] Jenkins and Van Kerm (2016) and Creedy and Gemmell (2019a; online appendix) also consider income changes in absolute terms, dx, as well as growth rates, dx/x.

the *inequality* of mobility within a given group is less straightforward, as it requires a visual comparison of (possibly multiple) slope changes across groups below p.

3 The TIM Curve

The empirical analysis in this paper focuses on the TIM and Re-ranking curve approaches to mobility measurement and illustration. This, and the next, section therefore summarise those approaches.

Jenkins and Lambert (1997) demonstrated that three important dimensions of cross-sectional poverty can be summarised by the following curve. Let x_i denote individual i's income, for $i = 1, ..., n$. For a specified poverty line, x_p, poverty gaps are defined by $g(x_i) = 0$ for $x_i > x_p$ and $g(x_i) = x_p - x_i$ for $x_i < x_p$. With incomes arranged in ascending order, plot $\frac{1}{n}\sum_{i=1}^{k} g(x_i)$ against $\frac{k}{n}$, for $k = 1, ..., n$. That is, the total cumulative poverty gap per capita is plotted against the associated proportion of people.

The curve conveniently displays the *incidence* of poverty (the headcount poverty measure), its *intensity* (the income gap, $x_p - x_i$), and its *inequality* (the dispersion of incomes below x_p). They therefore named the curve the 'three Is of poverty', or TIP, curve. The slope at any point is equal to the average poverty gap. A flattening of the curve therefore shows the extent to which the average poverty gap falls as income rises towards x_p. Thus, inequality among the poor is reflected in the curvature of the TIP curve. The curve becomes horizontal beyond x_p. Poverty is unambiguously higher where a TIP curve lies wholly above and to the left of an alternative TIP curve.

The TIP curve relates to poverty within a specified period of time over which income is measured. However, it is possible to define a related curve in the context of income growth between two periods. Creedy and Gemmell (2019a) define the 'three Is of mobility', or TIM, curve as follows. Define the logarithm of income, $y_i = \log x_i$, for individuals $i = 1, ..., n$. Hence $y_{i,t} - y_{i,t-1}$ is (approximately) person i's proportional change in income from period $t-1$ to t. With log incomes ranked in ascending order, plot $\frac{1}{n}\sum_{i=1}^{k}(y_{i,t} - y_{i,t-1})$ against $h = \frac{k}{n}$, for $k = 1, ..., n$.

Thus the TIM curve plots the cumulative proportional income change per capita against the corresponding proportion of individuals, h. The difference from the CIGP is that the measure of on the vertical axis is obtained by dividing by n rather than k.

This apparently small modification is important, since the properties of this alternative curve can more readily illustrate the three mobility characteristics of interest.

A TIM curve allows focus on the mobility of a particular group of low-income individuals: those with incomes below $x(h)$, for the proportion, h, of the population. In this framework h captures the *incidence* of the particular group of concern. Similarly, the *intensity* and *inequality* dimensions of mobility in terms of income growth are reflected in the shape of the TIM curve, by anology with the TIP curve.

The TIM curve can be specified more formally as follows, ignoring i subscripts for convenience. Suppose incomes are described by a continuous distribution where $H(x_t)$ and $F(y_t)$ denote respectively the distribution functions of income and log-income at time t, with population size, n. For incomes ranked in ascending order, the TIM curve plots the cumulative proportional income changes, $y_t - y_{t-1}$, per capita, denoted $M_{h,t}$, against the corresponding proportion of people, h, where:

$$h = F(y_{h,t-1}) \tag{4}$$

Thus $y_{h,t-1} = F^{-1}(h)$ is log-income corresponding to the h^{th} percentile, and the TIM curve plots $M_{h,t}$, given by:

$$M_{h,t} = \int_0^{y_{h,t-1}} (y_t - y_{t-1})\, dF(y_{t-1}) \tag{5}$$

against h.

Let μ_t denote the arithmetic mean of log-income (that is, the logarithm of the geometric mean, G_t, of income, x_t. Equation (5) can be written as:

$$M_{h,t} = \int_0^{y_{h,t-1}} \left\{ (y_t - \mu_t) - \left(y_{t-1} - \mu_{t-1}\right) \right\} dF(y_{t-1}) + \left(\mu_t - \mu_{t-1}\right) F(y_{h,t-1}) \tag{6}$$

The term, $y_t - \mu_t$ is equal to $\log(x_t/G_t)$. Hence $(y_t - \mu_t) - \left(y_{t-1} - \mu_{t-1}\right)$ is the proportional change in *relative* income. Thus, $M_{h,t}$ consists of the cumulative proportional change in income *relative* to the geometric mean, *plus* a component that depends only on the proportional change in geometric mean income.

Let g denote the proportional change in geometric mean, $\mu_t - \mu_{t-1}$, and suppose the proportional change in relative income depends on income in $t-1$, so that $(y_t - \mu_t) - \left(y_{t-1} - \mu_{t-1}\right)$ can be written as the function, $g^*(y_{t-1})$. Then (6) can be expressed as:

$$M_{h,t} = \int_0^{y_{h,t-1}} g^*(y_{t-1})\, dF(y_{t-1}) + gh \tag{7}$$

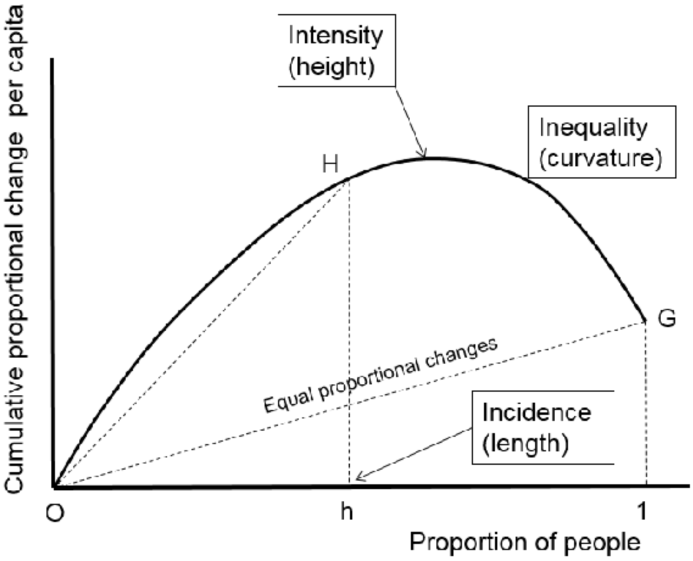

Figure 1: A TIM Curve

If all individuals receive exactly the same relative income change, then relative positions are unchanged and $g^*(y_{t-1}) = 0$ for all y_{t-1}. Hence, $M_{h,t}$ plotted against h is simply a straight line through the origin with a slope of g. This means that the extent to which it is equalising or disequalising over any range of the income distribution can be seen immediately by the extent to which the TIM curve deviates from a straight line, which in turn depends on the properties of $g^*(y_{t-1})$.

A hypothetical example of a TIM curve is shown in Figure 1, with $h = k/n$ on the horizontal axis. This reflects a situation in which relatively lower-income individuals receive proportional income increases which are greater than that of average (geometric mean) income. Hence the TIM curve, OHG, lies wholly above the straight line OG.

If all incomes increase by the same proportion, the TIM curve is the straight line OG. The height, G, indicates the average growth rate of the population as a whole, with the height, H, indicating the average growth rate for those below $x(h)$. Furthermore, inequality is reflected in the degree of curvature. For example, the curvature of the arc OH relative to the straight line OH indicates that lower income individuals have higher (more unequal) growth than those individuals to the left of, but closer to, h.

Suppose interest is focussed on those below the h^{th} percentile, indicated in Figure

1. There is less 'inequality of mobility' within the group below h, shown by the fact that the TIM curve from O to H is closer to a straight line than the complete curve OHG.[12] The TIM curve also shows that the income growth of those below h is larger than that of the population as a whole. The average growth rate among the poor (the intensity of their growth) is given by the height H.

If it is preferred to assess mobility from *relative* income growth rates, some normalisation of the TIM curves is required. For example, comparing the income mobility experienced across different periods, the mean income growth rate, g, is likely to vary across periods, such that the height of point G in Figure 1 differs. This can make companions of the degree of 'inequality of mobility', the third 'I', across periods difficult. In this case equivalent 'normalised TIM' curves, or 'nTIM' curves, can readily be obtained where each TIM is normalised by the sample average growth rate for each period. With normalisation, $M_{h,t}$ reaches a value of 1 at $h = 1$, though $M_{h,t}$ values can exceed 1 at lower values of h, as illustrated in Figure 1. This normalisation allows the degree of concavity or convexity of each TIM curve to be directly compared.

[12] There is a potential ambiguity in the use of the term 'inequality' here since the TIP curve refers to a cross-sectional distribution whereas the TIM curve refers to income changes. To avoid confusion over nomenclature, when referring to the 'inequality dimension' of mobility (one of the three 'I's), the term 'interpersonal dispersion' of mobility is perhaps preferable.

4 Positional Mobility

An alternative mobility concept is based on the idea of mobility as positional change, rather than relative income growth.[13] This section focuses on income re-ranking measures first proposed and illustrated by Creedy and Gemmell (2019a). Individuals can obviously move to higher or lower positions, so the explicit treatment of the *direction* of change is necessary. In the following, individuals are ranked in ascending order of initial incomes, $x_{i,0}$, so that $i = 1, ..., n$ orders individuals from the lowest to the highest income. The initial period is 0, and initial ranks are $R_{i,0} = i$. First, a choice must be made regarding *whose* mobility to be included. Here, concentration is on a subset of individuals, $k \leq n$, with the lowest initial incomes. Second, it is necessary to decide whether negative re-ranking (dropping down the ranking) is treated symmetrically with upward (positive) movement.

Let $\Delta R_i = R_{i,1} - R_{i,0} = R_{i,1} - i$ denote the change in the rank of the person who initially has rank, i. Three options are possible, depending on how negative re-ranking is treated. First, negative re-ranking can be treated symmetrically with positive re-ranking such that positional mobility is defined in *net* terms, that is, positive changes in rank net of any negative changes within group $i = 1, ..., k$.[14] This is referred to as '*net* re-ranking'. Secondly, negative movements could be ignored, which simply involves setting $\Delta R_i = 0$ when $\Delta R_i < 0$: this is referred to as '*positive* re-ranking'. Thirdly, re-ranking may be measured in *absolute* terms in which all re-ranking is positive: this is referred to as '*absolute* re-ranking'. The choice among these three measures depends on the question of interest. For example, if interest is focussed on those below the poverty line *as a group*, then it may be desirable to balance any upward mobility by some of those in poverty with downward mobility of others in poverty, to gain an indication of the net experience of the group. This suggests a focus on net mobility. If movement *per se* is the mobility concept of interest, a non-directional measure such as absolute

[13] D'Agostino and Dardanoni (2009) and Cowell and Flachaire (2016) have sought to re-define and clarify various rank-related mobility concepts and measures. Cowell and Flachair (2016) propose a 'superclass' of rank-based measures. They stress the importance of separating the evaluation of an individual's positional 'status' from movements between positions, where measurement of the latter uses distance concepts. However, neither study offers graphical devices to illustrate the measures.

[14] If individual changes in rank are simply aggregated to obtain an aggregate mobility index, then a change in rank of 50 places by one individual is treated symetrically as 50 individuals each changing one ranking place.

re-ranking is relevant. Positive re-ranking quantifies only those who are moving up, a common metric when assessing the persistence of low income or poverty.[15]

The three re-ranking indices individual, i, can be defined formally (where pos = positive; abs = absolute) as $M_i^{net} = \Delta R_i$, $M_i^{pos} = \Delta R_i|_{\Delta R_i \geqslant 0}$, and $M_i^{abs} = |\Delta R_i|$. Cumulated across the k lowest income individuals in period 0, the corresponding aggregate re-ranking indices are:[16]

$$M_k^{net} = \sum_{i=1}^{k} M_i^{net} = \sum_{i=1}^{k} (R_{i,1} - R_{i,0}) \tag{8}$$

$$M_k^{pos} = \sum_{i=1}^{k} M_i^{pos} = \sum_{i=1}^{k} (R_{i,1} - R_{i,0}) \quad \text{for} \quad \Delta R_i \geqslant 0 \tag{9}$$

$$M_k^{abs} = \sum_{i=1}^{k} M_i^{abs} = \sum_{i=1}^{k} |R_{i,1} - R_{i,0}| \tag{10}$$

To examine the 'three Is' of positional mobility, using (8), (9) and (10), one approach would be to plot the value of the relevant M_k index against the cumulative fraction of the population, $h = k/n$. However, there are two difficulties with the indices in (8) to (10). First, they are not scale independent, since they depend on k and hence population size, as more re-ranking is possible in larger populations. One solution would be to scale the three M_k indices by n. However, as is shown below, a slightly different rescaling, by $(n/2)^2$, yields normalised values, m_k, that lie between zero and one (or zero and two for positive re-ranking). These may be plotted against $0 \leqslant h \leqslant 1$.

Secondly, an individual's *opportunity* for re-ranking is partly determined by the initial position: someone among the lowest ranks has less opportunity to move down, other things equal, than someone higher up, and *vice versa*. It is therefore useful to consider the maximum re-ranking possible for each individual; actual re-ranking may then be compared with these maximum values for any given h.

[15] On poverty persistence, see Creedy and Gemmell (2018).

[16] The absolute re-ranking case may be thought of as describing overall positional change within the relevant income range. Over short periods this is often described as volatility, or 'income risk', with a presumption that, *ceteris paribus*, less risk is preferable to more risk. Over longer time periods it may be regarded as describing the flexibility of the income distribution. This has less-clear welfare associations, although greater long-term mobility is often characterised as implying less-intrenched social inequalities.

Consider first the maximum re-ranking and, to simplify the exposition, consider a population of $n = 100$ individuals, each with a different income level; hence each integer, $i = 1, ..., n$, represents a percentile. They are ranked in period 0, $R_{i,0} = 1 \ldots 100$, representing the lowest to the highest incomes. Two polar cases are the maximum and minimum degrees of mobility possible. The former is defined here as a complete ranking reversal, $\Delta R_i(\max)$, such that in period 1, $R_{i,1}$ involves a lowest to highest ranking of $R_{i,1}(\max) = n + 1 - R_{i,0} = 100, \ldots, 1$.[17] Maximum re-ranking implies:

$$M_i(\max) = \Delta R_i(\max) = R_{i,1}(\max) - R_{i,0} = n + 1 - 2R_{i,0} \tag{11}$$

which, for large n, can be approximated by $n - 2R_{i,0}$. Where it is desired to measure the extent of re-ranking of the subset of individuals, $k \leq n$, with the lowest incomes, the cumulative maximum re-ranking index for the net mobility case , $M_{k}^{net}(\max)$, is:

$$M_{k}^{net}(\max) = \sum_{i=1}^{k} M_{i}^{net}(\max) = \sum_{i=1}^{k}(n + 1 - 2R_{i,0}) \tag{12}$$

Using the sum of an arithmetic progression, whereby $\sum_{i=1}^{k} R_{i,0} = 1 + 2 + \ldots + k = k(k+1)/2$, equation (12) becomes:

$$M_{k}^{net}(\max) = \sum_{i=1}^{k}(n + 1 - 2R_{i,0}) = k(n+1) - k(k+1) = k(n-k) \tag{13}$$

Hence, in the $n = 100$ example above, if interest focuses only on the poorest individual ($k = 1$), maximum net re-ranking is given by $M_{k}^{net}(\max) = (100-1) = 99$; when $k = 2$, $M_{k}^{net}(\max) = 2(100-2) = 196$; and so on. More generally, since maximum re-ranking (complete ranking reversal) involves all those below the median individual changing positions with those above the median, it follows from (13) that the maximum value of $M_{k}^{net}(\max)$ as k increases is obtained for $k = n/2$, yielding $M_{k}^{net}(\max) = (n/2)^2$.[18]

This measure therefore serves to highlight the scale dependence of both M_{k}^{net} and $M_{k}^{net}(\max)$: larger populations imply larger values of both indices. These could be

[17] Jantti and Jenkins (2015; pp. 8-9) proposed that the relevant comparator should be defined as when the change in an individual's position is purely random. That is, 'maximum' mobility involves independence from initial positions, rather than complete reversals. They reject the use of 'maximum' when mobility is based on origin independence.

[18] Strictly, for small n, the median individual is $k = (n+1)/2$, and $M_{k}^{net}(\max)$ is given by $(n+1)(n-1)/4$.

'normalised' to create a form of per capita index by dividing by n^2 such that, from (13), the index becomes: $m_k^{net}(\max) = h(1-h)$. The maximum value would be reached at $h = 0.5$, where $m_k^{net}(\max) = 0.25$. However, to get an index with a maximum value of 1 (at $k = n/2$) , it is preferable to divide by $(n/2)^2$. That is:

$$m_k^{net}(\max) = 4M_k^{net}(\max)/n^2 \tag{14}$$

Using (13):

$$m_k^{net}(\max) = 4h(1-h) \tag{15}$$

A similar exercise for positive re-ranking, $M_k^{pos}(\max)$, shows that the value of $M_k^{pos}(\max)$ also reaches a maximum as k increases of $M_k^{pos}(\max) = n^2/4$ when $k = n/2$, since all individuals below $n/2$ experience positive re-ranking in this (maximum) case. However, above $k = n/2$, as more above-median individuals are included within k, their re-rankings are now given by $\Delta R_i = 0$, such that the cumulative index, $M_k^{pos}(\max)$, remains unchanged as $k \longrightarrow n$. Thus a similarly rescaled $m_k^{pos}(\max)$ may be defined analogously to (14) to yield a positive re-ranking index where $0 \leqslant m_k^{pos}(\max) \leqslant 1$.

Finally, for the absolute re-ranking case in (10), $M_k^{abs}(\max)$, this increases as k increases from $k = 1$ to $k = n/2$ to reach $M_k^{abs}(\max) = (n/2)^2$. However, this is a point of inflection rather than a maximum, since inclusion of the absolute value of above-median individuals' re-ranking in $M_k^{abs}(\max)$, ensures that $M_k^{abs}(\max)$ continues to increase for $k > n/2$, reaching $M_k^{abs}(\max) = n^2/2$ at $k = n$. Hence, an absolute re-ranking index $m_k^{abs}(\max)$ obtained by rescaling by $(n/2)^2$ lies between zero and two.

Finally, to compare actual and maximum re-ranking mobility, the expressions for actual mobility in (8) to (10) can be similarly rescaled or normalised by $(n/2)^2$ to obtain actual aggregate re-ranking mobility expressions, m_k^{net}, m_k^{pos}, and m_k^{abs}, given in each case by:

$$m_k = 4M_k/n^2 \tag{16}$$

Thus, $0 \leqslant m_k^{net}, m_k^{pos} \leqslant 1$ and $0 \leqslant m_k^{abs} \leqslant 2$. This suggests a convenient illustrative device for positional mobility is a cumulative re-ranking curve that plots m_k against h.

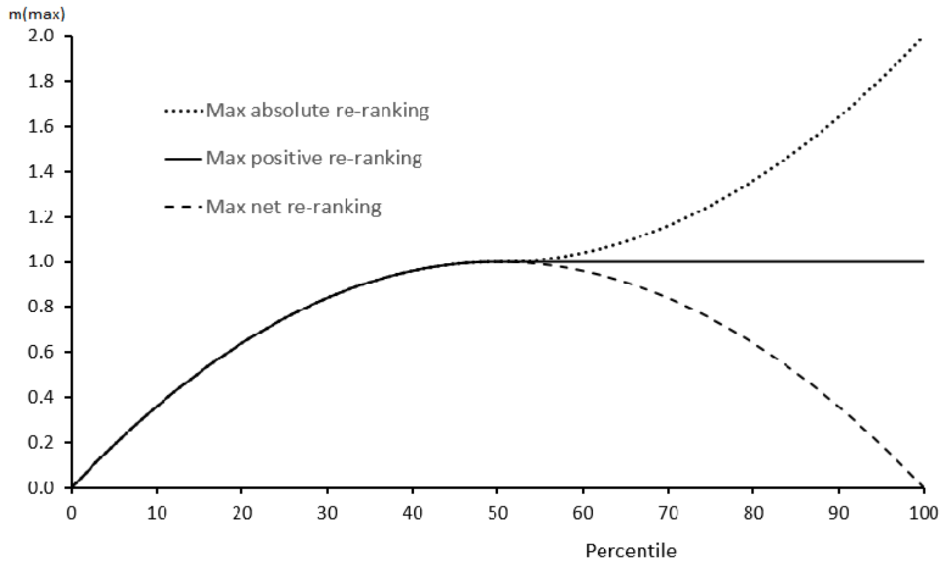

Figure 2: Maximum Re-ranking

4.1 Maximum Re-Ranking profiles

Profiles for the three (rescaled) maximum re-ranking cases discussed above, $m_k^{net}(\max)$, $m_k^{pos}(\max)$, and $m_k^{abs}(\max)$ are plotted against $h = k/n$ in Figure 2. This shows the distinct non-linear shape of the maximum profiles, whichever definition of positional mobility is adopted – net, positive or absolute. As expected, the net re-ranking profile displays a parabolic shape which differentiation of (15) reveals has a slope of $4(1-2h)$; hence equals zero at $h = 0.5$ (the 50^{th} percentile), thereafter declining symmetrically to a slope of -4 at $h = 1$. The equivalent positive re-ranking profile also reaches a maximum at the 50^{th} percentile but remains constant thereafter, while the absolute re-ranking profile displays a sigmoid shape, reaching a local point of inflection where $m_k^{abs}(\max) = 1$ at the 50^{th} percentile, but then rising at an increasing rate till $m_k^{abs}(\max) = 2$ at $h = 1$.

The maximum re-ranking indices in Figure 2 are invariant to population size, but vary with the population percentile, of interest, h. Thus, the *scope* for a given degree of re-ranking also varies with h. A natural index of interest therefore is the ratio of actual to maximum mobility at each percentile, h. This is referred to below as the 're-ranking ratio', RRR_k, and can be calculated for net, positive and absolute re-ranking.

For example, the net re-ranking case is given by:

$$RRR_k^{net} = \frac{m_k^{net}}{m_k^{net}(\max)} = \frac{M_k^{net}}{M_k^{net}(\max)} \tag{17}$$

where the numerator and denominator are given respectively by (16) and (14), or by (8) and (13). This ratio can also be plotted against h to identify how the extent of mobility changes by percentile of the population relative to the maximum possible *for that percentile*. Recognising these differences in maximum re-ranking is important when interpreting differences in actual re-ranking for different values of h. In particular, a smaller value of m_k^{net} at $h = 0.1$, compared to m_k^{net} at $h = 0.3$, for example, may be partly or entirely due the fact that individuals up to $h = 0.1$ cannot achieve the higher m_k^{net} observed at $h = 0.3$, even in the absence of other constraints on re-ranking mobility.

5 The Longitudinal Dataset

This section summarises the longitudinal dataset used below: a detailed description and explanation of its construction is given in Alinaghi *et al.* (2020). The dataset has been made possible by the improved availability of anonymised administrative register data, such as from individuals' tax records, in New Zealand's Integrated Data Infrastructure (IDI). This has facilitated the construction of longitudinal data through the matching of income records for individuals over time. These data sources provide several advantages compared to surveys, such as very large sample sizes, improved coverage of top incomes, avoidance of survey respondent dropout or attrition, and less measurement error. The data used in this paper provides the most comprehensive information to date on NZ taxpayers' incomes, suitable for inequality and mobility analysis.

A number of administrative datasets within the IDI were merged to form the final dataset used here. The primary database covers the Inland Revenue individual taxpayer population, containing detailed tax return and PAYE information such as wage and salary earnings, self-employment income, pensions, and capital income. Socioeconomic variables such as gender, age, ethnicity and highest educational qualification were then added to the primary dataset. From a population of around 5.4 million taxpayer observations for whom there is taxable income information in the IDI for at least one year of data over the 18 years 2000 to 2017, a sub-sample of around 1.5 million individuals is available with income data for all 18 years.

For the present exercise it was decided to start with the income distribution in 2002 rather than 2000, thus covering 16 years of income data, or 15 years of income growth for all individuals. This reduces the sample size slightly, to around 1.450 million individuals, but avoids potential distortions associated with the 2000-2001 years when reforms to the top personal income tax rate are known to have caused annual taxpayer incomes, especially towards the top of the income distribution, to fluctuate temporarily; see Creedy *et al.* (2021).

Table 1 shows some decompositions of the total taxpayer population with annual data over the 2002 to 2017 period, by gender, age, ethnicity and highest educational qualifications. This indicates that the gender composition is close to 50:50 between males and females. Māori and Pasifika represent around 14 per cent and 4 per cent respectively of all individuals. Other ethnicities recorded in the dataset include European, Asian, Middle Eastern/Latin American/African and 'Other' (miscellaneous) represent the remaining 86 per cent.[19]

For longitudinal data covering a large number of years, defining the working age group is not straightforward. The table shows outcomes using two definitions. The first case defines working age individuals as those aged 20 to 64 in 2002.[20] This may be regarded as most suitable for mobility measured over 1 year, for example, 2002 to 2003. The second working age definition considers only those aged 20-64 in all years 2002 to 2017, hence including only those aged 20-49 in 2002. These two definitions yield working age sub-groups of 86 per cent of the total sample (1.248 million) and 63 per cent of the total (0.912 million) respectively. TIM and nTIM curve results reported below use the second working age definition, but both definitions generate very similar curves.

For educational qualification decompositions in Table 1, data on highest educational qualification are constructed such that individuals are assigned to a category according to their highest qualification obtained in any year during the 2002 to 2017 period. For example, an individual obtaining a university degree in 2005 is allocated to this category throughout the period examined. This avoids changes in sub-sample sizes for each qualification category during the period, and reflects the interest here in an income decomposition based on an individual's educational capability or potential (as demonstrated by their highest qualification) rather than distinguishing incomes pre- and post-qualification.[21]

[19]In the 2018 New Zealand census, out of a total population of 4,699,755 individuals, ethnicity percentages were as follows: European (70), Māori (17), Pasifika (8), Asian (15), MELAA (Middle Eastern, Latin American, and African) (1), Others (1). These percentages add to more than 100 percent because individuals are able to specify more than one ethnicity. In the dataset used here a single 'prioritised ethnicity variable' has been created by assigning ethnicity to each individual according to the following priority ordering: Māori, Pacific Peoples, Asian, European, MELAA, and Other. For example, an individual is classified as Māori, if their ethnic code in one of the three data sources is Māori. This process is repeated for other ethnic groups in order; see Alinaghi *et al.* (2020, p.11-12) for further details.

[20]Of course actual working ages differ across individuals with many working, especially part-time, before age 20 and after age 64. The relatively restrictive working age definition of 20-64 aims to focus attention on those most likely to be permanently attached to the workforce, after any post-school education and prior to receipt of New Zealand Superannuation.

[21]Some individuals may go on to obtain an additional, higher qualification in the years after the final year of the dataset in 2017, which obviously cannot be captured here.

Around 20 per cent of the total have no qualifications (250,140 individuals). This is similar to those with university degrees (18 per cent), while individuals with 'school' and 'post-school' qualifications represent around 36 and 26 per cent of the total respectively. 'Post-school' qualifications include diplomas and other non-degree qualifications from higher education institutions such as technical colleges and Wānanga.

Table 1: Sample Sizes by Decomposition

	Sample size		Sample size
Gender:		Ethnicity:	
Male	736,371	Māori	200,451
Female	711,384	Pasifika	64,692
		Non-Māori, non-Pas.	1,182,612
Total	1,447,755	Total	1,447,755
Age:		Educational Qualifications:**	
Working*	1,248,510	None	250,140
Non-working	214,239	School	457,917
Working§	912,018	Post-school	325,521
Non-working	535,737	University	222,543
Total	1,447,755	Total	1,256,121

*Ages 20-64 in 2002. §Ages 20-64 in all years, 2002-2017.

**Educational sub-totals sum to smaller total due to missing qualifications data for some individuals.

Using these longitudinal data to construct TIM and nTIM curves for various periods during 2002 to 2017, it would of course be possible to use larger sample sizes for shorter periods since attrition tends to reduce sample sizes the longer the time period considered. However, to aid comparability of results for different periods, the curves presented in section 6 are each based on the same longitudinal sample summarised in Table 1.

6 TIM and nTIM Curves for New Zealand

This section illustrates the TIM and nTIM curves described in section 3 based on the dataset described in the previous section. To consider mobility over various time periods, TIM and nTIM curves are constructed for the same taxpayers over different time periods: 1 year (2002-2003); 5 years (2002-2007); 10 years (2002-2012); and 15 years (2002-2017).[22] Subsection 6.1 first focuses on income mobility measured across all individuals, while subsequent subsections examine decompositions by gender, age, ethnicity, and educational qualifications.

6.1 All Taxpayers

Figure 3 shows TIM and nTIM curves, in upper and lower panels respectively, corresponding to each of the four periods. In each case, as in all diagrams below, individuals are ranked by their 2002 incomes, with percentiles of the income distribution, h, in 2002 on the horizontal axis. Cumulative growth rates per capita, $M_{h,t}$, measured over the entire period, are shown on the vertical axis. As a result, the right-hand end of each TIM curve, which represents the average growth rate across all individuals over the whole period (1, 5, 10, 15 years) shifts vertically as the period considered is extended. For example, the four TIM curves in the top panel of Figure 3 show that the average cumulative growth rate of taxable income per capita across the full sample ($h = 1$; the 100^{th} percentile) was around 0.05 (5 per cent) over 2002-2003; 0.25 over 5 years, 2002-2007; 0.4 over 10 years, 2002-2012; and 0.47 over 15 years, 2002-2017.

Although the straight 'lines of equal mobility' from the origin to the end point of each TIM curve are not shown in the diagram (to facilitate visual clarity) it is immediately clear that all four period TIM curves display concave properties. That is, the average income growth rates experienced by those in the lower percentiles of the initial income distribution exceed the equivalent growth rates that would have been observed if those same individuals had experienced income growth equal to that of all taxpayers combined. This could be described as 'pro-poor' mobility since, in all four

[22] Since this dataset includes some individuals on very low incomes (such as small part-time earnings of children, or small capital incomes of non-earners), TIM and nTIM curves were constructed for the full sample, and also for samples restricted to those individuals with incomes in any year above \$1,000, \$5,000, \$10,000 and \$20,000. Results were found not to be sensitive to those exclusions; they are therefore reported below for the full sample

periods, the *relative* growth of those initially on lower incomes exceeds that of those initially on higher incomes. Given well-known issues around defining which sub-groups are included in the 'poor' category, the discussion below uses the term 'progressive' income growth to indicate income growth which is greater among individuals initially with relatively lower incomes than those initially with higher incomes (and *vice versa* for 'regressive' growth).[23]

The progressive aspect observed with the TIM curves in Figure 3 holds even over the longest period examined of 15 years, 2002 to 2017. As Creedy and Gemmell (2019b) stress when comparing income inequality measured using cross-sectional and longitudinal data, this progressive property of income mobility is often obscured when examining year to year changes in annual cross-section snapshots of income inequality. However, it appears to be a robust property of mobility when examining the income progress of the *same people* over time.

The lower panel of Figure 3 has been constructed to illustrate the extent of interpersonal dispersion (inequality) of mobility for the sample as a whole, via comparisons of the concavity of the four nTIM curves relative to the common straight line representing equality of mobility.[24] This clarifies differences in the extent of progressive growth between periods, and highlights the fact that it tends to be most pronounced over shorter periods: the nTIM curve for 2002-2003 lies wholly above the curve for 2002-2007, which generally lies above the 2002-2012 and 2002-2017 curves. The latter two curves are harder to distinguish, tending to overlay and cross each other, suggesting that the progressive mobility patterns observed in the 10- and 15-year nTIMs may approximate a more sustained long-term characteristic.[25]

As mentioned in section 1, it is important to be cautious when drawing normative conclusions from these nTIM results. In addition to the well-known issues around defining and measuring social welfare, and value judgements implicit in inter-personal comparisons more generally, it is unclear how far income variability should be regarded

[23]The terms progressive income mobility does not necessarily imply a systematic movement towards the arithmetic mean log-income (geometric mean), merely inequality-reducing mobility.

[24]If interest is focused on the interpersonal distribution of mobility for a particular income group, such as the poorest half of the sample, $h = 0.5$, then each TIM curve can be 're-normalised' using the average income growth rate for this group.

[25]This is also reflected in a flattening of profiles of an inequality measure as the accounting period increases: see Creedy *et al.* (2021).

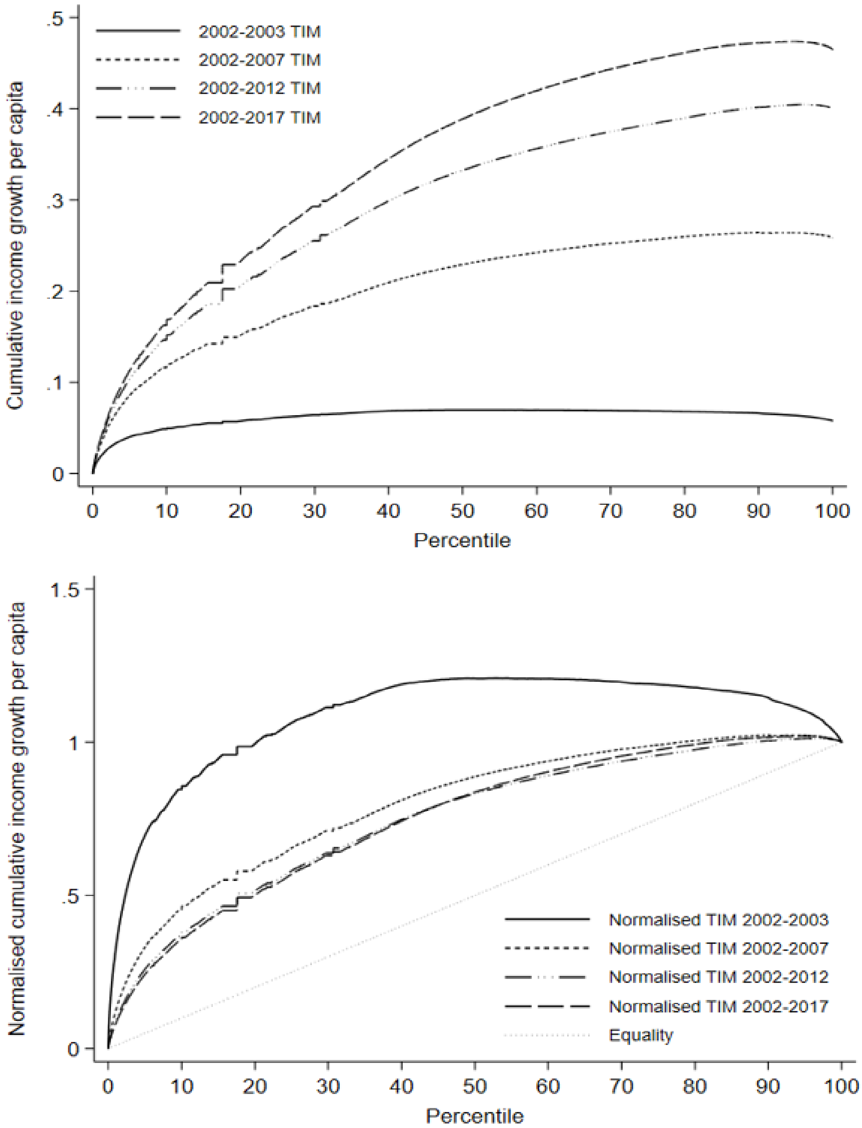

Figure 3: TIM and Normalised TIM Curves: All Individuals

as welfare enhancing or retarding. For individuals with decreasing marginal utility of income, a smooth income stream would be regarded as preferable to a volatile one. Hence, the high income volatility implicit in the 1-year nTIM result could be thought of as undesirable. However, longer-term movements in incomes, such that those initially on lower incomes grow faster than their higher-income equivalents are often argued to represent 'improvements'. The point at which undesirable short-term income volatility transforms into desirable longer-term social mobility is clearly important to understand but not straightforward to establish.

6.2 TIMs by Gender and Age

Patterns of income mobility might be expected to differ by gender and by age. For example, women on average tend to earn less than men in part due to lower attachment to the labour force in general and greater use of part-time working in particular. Thus, women and men are not randomly distributed across the combined income distribution such that income mobility outcomes may also differ by gender.

By age, an obvious difference concerns workers versus retirees. Retirees are known to be generally less mobile geographically and by occupation, with implications for income mobility. In addition many retirees in New Zealand have relatively fixed, and relatively low, incomes associated with the universal property of New Zealand Superannuation. Measures of overall income mobility of individuals of working age may therefore look somewhat different than when non-working individuals are included. This subsection examines TIM and nTIM curves separately for males and females and also for individuals of working age (using the definition in section 5 of those aged 20-64 years during 2002 to 2017).

Examining separate TIM and nTIM curve for males and females reveals that both are very similar to those represented in Figure 3 for all individuals. Across-gender differences are most easily observed using nTIM curves as shown in Figure 4, for the same four periods. The Figure suggests that income mobility was slightly more progressive for females than males over the short 2002 to 2003 period: lower-income females did relatively better than lower-income males.

However, mobility over 5 years, 2002 to 2007, shows little difference, although the female nTIM remains above the male nTIM, while over 10 and 15 years, the two nTIMs become almost indistinguishable. Indeed after 15 years, the female nTIM may even be slightly below the male equivalent from around the 30^{th} percentile upwards. This implies slightly less progressive mobility for females than males when higher-income individuals of both genders are included in the comparison.

Considering mobility of only working age individuals in Figure 5 reveals a similar pattern to that shown in Figure 3 for all individuals. Comparing the TIM curves in the two Figures shows that both display the same approximate shape for equivalent periods, 2002-2003, 2002-2007 and so on. However, over longer periods the income growth rates for working age individuals tend to exceed those for all individuals combined. For example, for the longest period, 2002 to 2017, average income growth across the working age population is around 0.6 (60 per cent over 15 years), whereas it is around 0.47 when non-working individuals are included. However, nTIM curves for both groups display similar tendencies for growth to be progressive, but for this to decrease as the period length is extended from 1 to 15 years.

6.3 TIMs by Ethnicity and Education

Patterns of income mobility may differ depending on the ethnic groups one specified or highest qualifications achieved. This section examines the potential differences by looking at nTIM curves for groups distinguished by these characteristics.

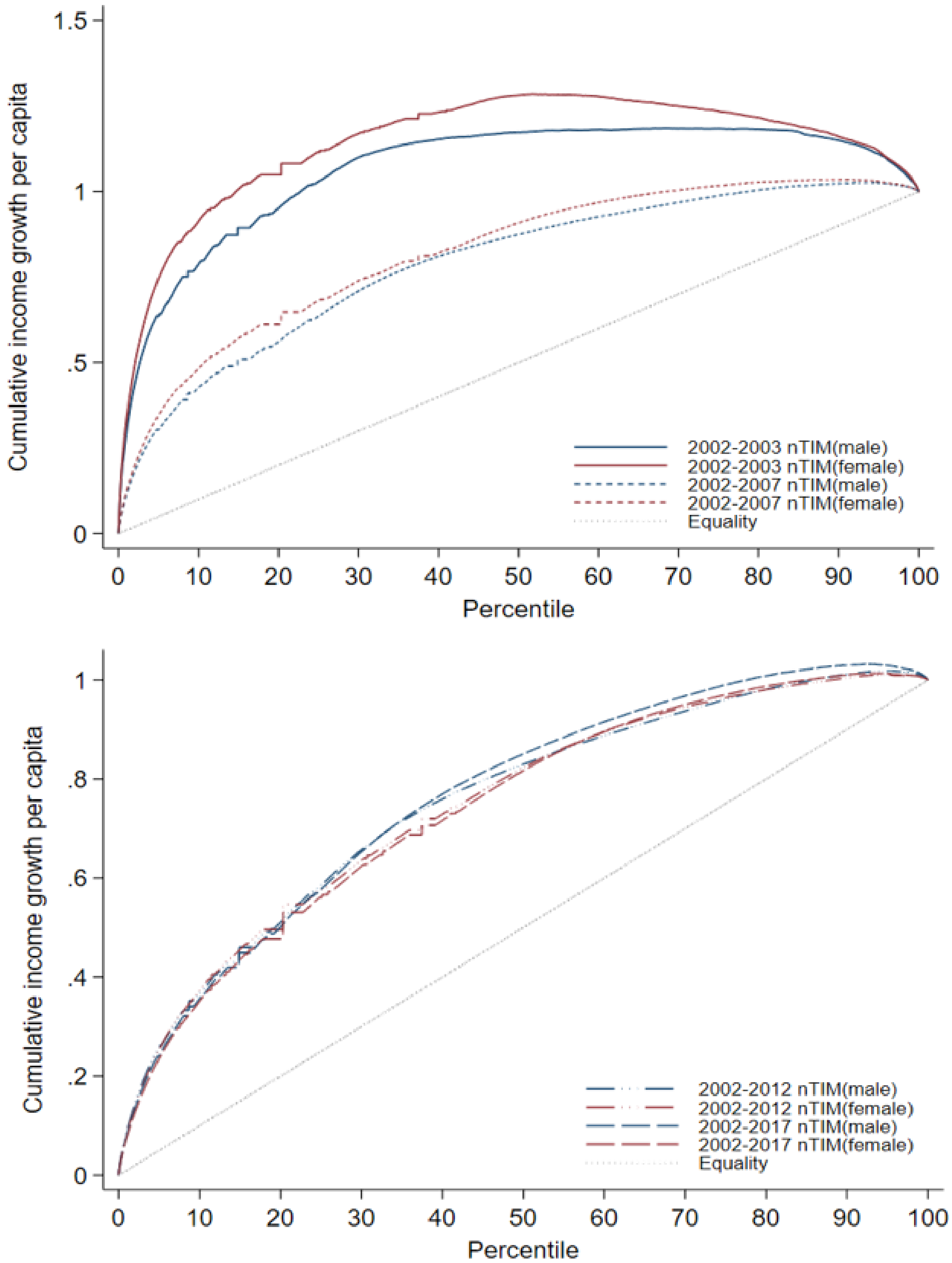

Figure 4: Comparing nTIM Curves for Males and Females

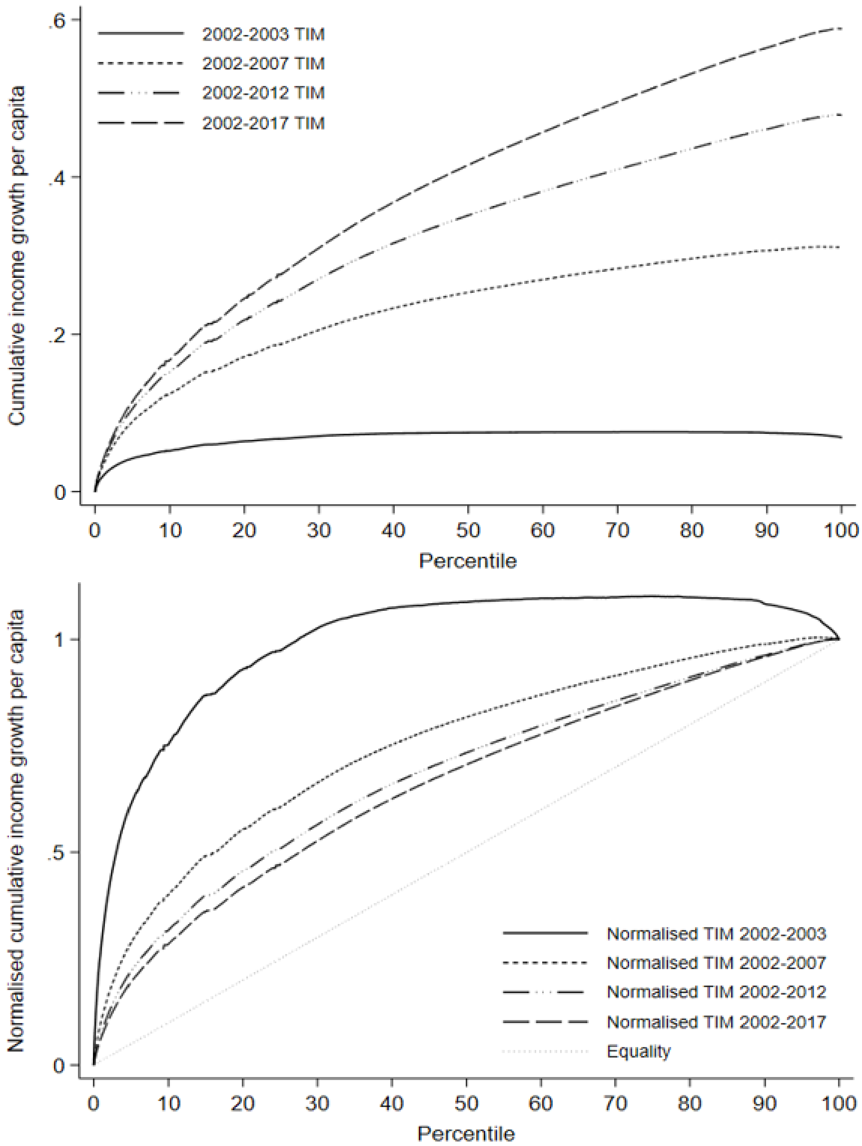

Figure 5: TIM and nTIM Curves for Working Age Individuals

Figure 6 illustrates nTIM curves for three ethnic groups, Māori, Māori and Pasifika, and all others combined (non-Māori, non-Pasifika). The normalised TIM curves for the first two groups, Māori and Māori and Pasifika, display similar tendencies for income growth which decrease as the period length is extended. For the last group, all others combined (non-Māori, non-Pasifika), on the other hand, the curves for 10 years (2002-2012) and 15 years (2002-2017) are almost indistinguishable until around 50^{th} percentile but then the curve for 2002-2017 lies wholly above the curve for 2002-2012. This is similar to that shown in Figure 3 for all individuals. Figure 7 shows four nTIM curves corresponding to four time periods decomposed by four educational qualifications. It is perhaps not surprising that the income profiles of individuals with school qualifications appears to be similar to those of post-school qualifications.

However, for those with no educational qualifications, the two nTIMs over 10 and 15 years become almost indistinguishable for up to about 50^{th} percentile but then the nTIM for 15 years period remains above the 10 years. This implies slightly more progressive mobility over the longer period. In the case of individuals with university degree, the Figure suggests that over the long period there is slightly lower upward mobility among the lower percentiles. This becomes reverse after around 30^{th} percentile upwards.

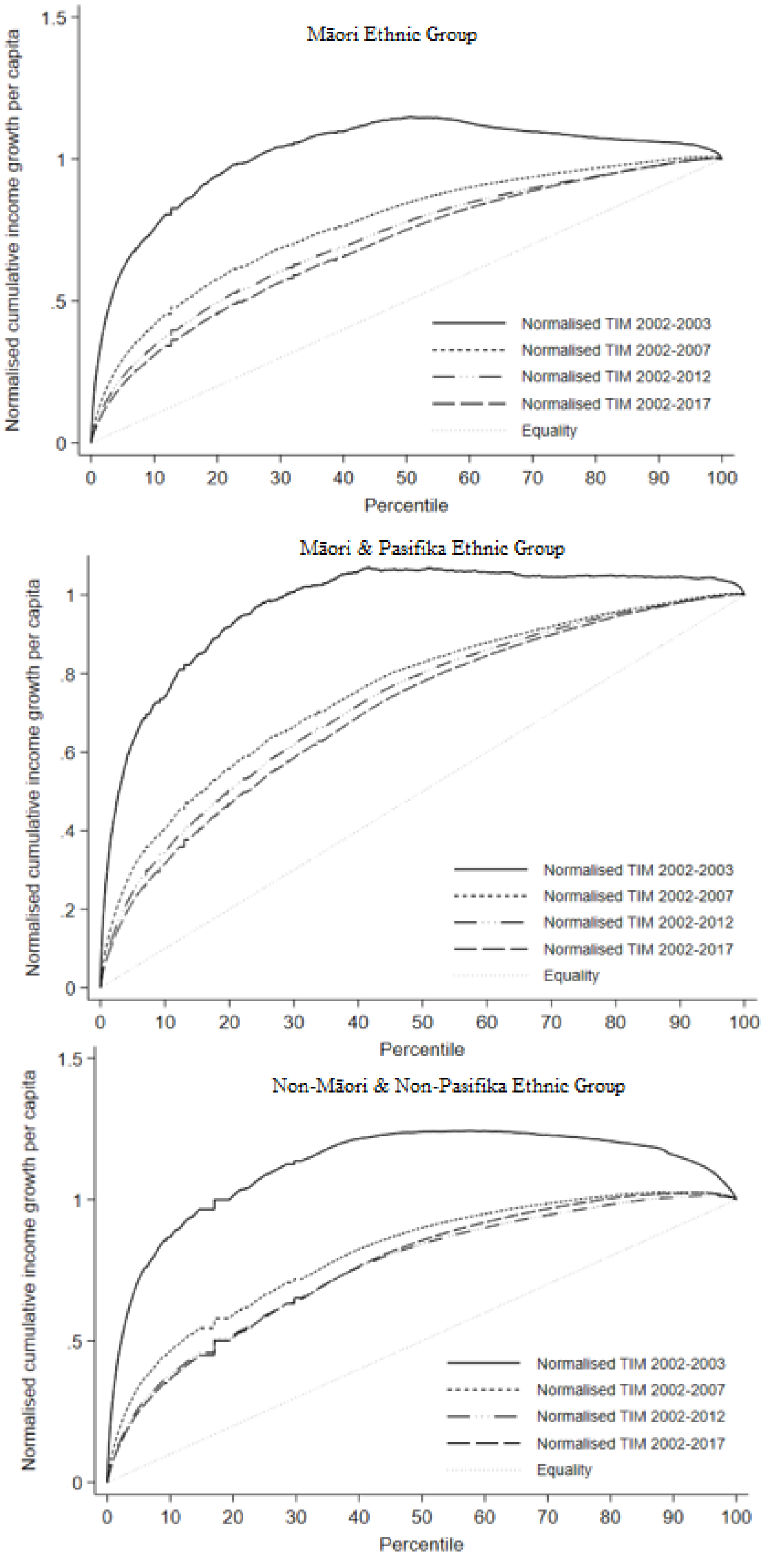

Figure 6: Comparing nTIM Curves for Ethnic Groups

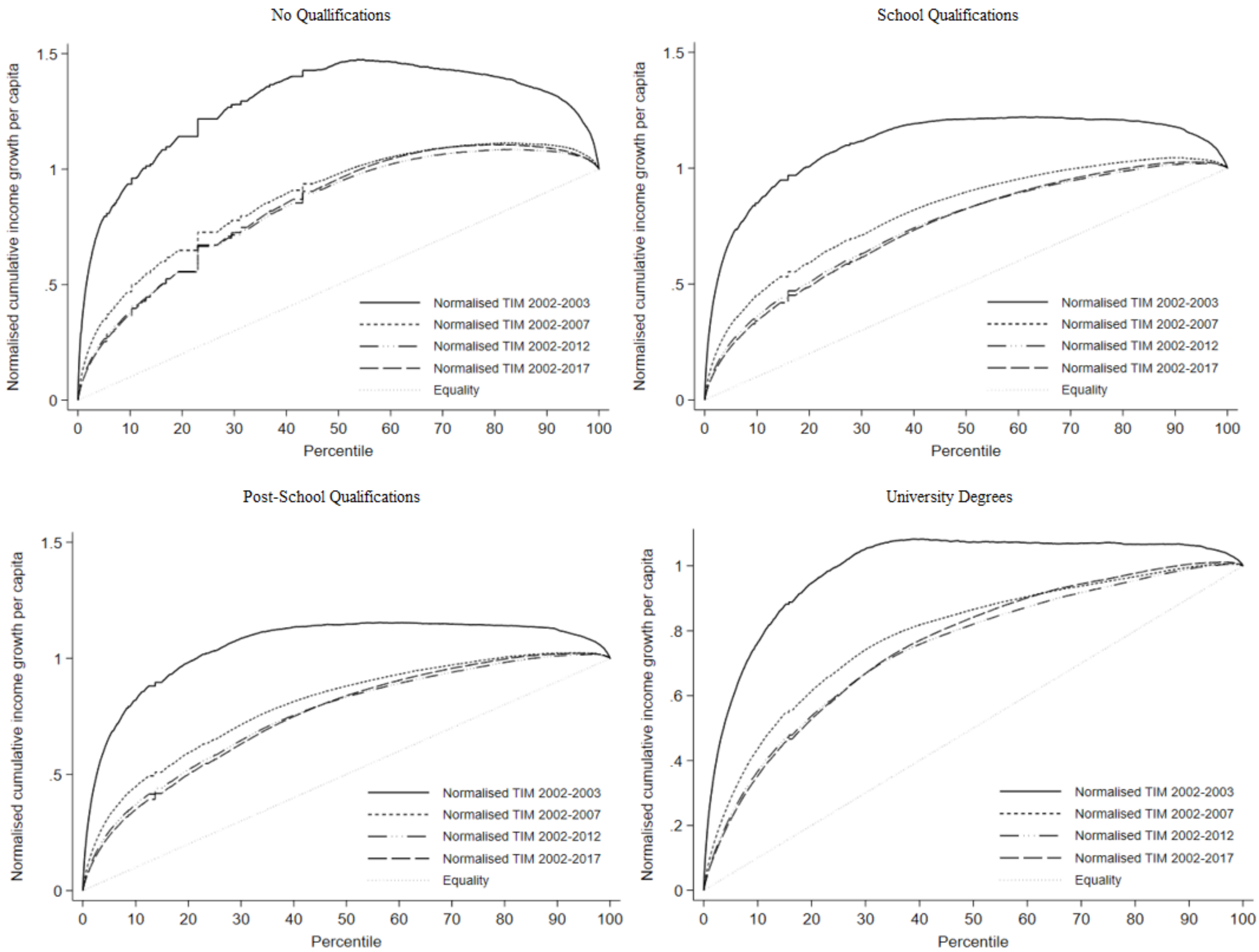

Figure 7: Comparing nTIM Curves for Educational Qualifications

7 Re-ranking Mobility in New Zealand

This section reports the positional mobility measures, described in Section 4, for the same New Zealand income data, to assess both the extent of observed positional mobility and its incidence, intensity and interpersonal dimensions. This is illustrated first by plotting the re-ranking measures m_k^{pos}, m_k^{net} and m_k^{abs} against h, analogous to the $m_k(\max)$ profiles in Figure 2.

To save space, in Figure 8 these are shown for the short 5-year period, 2002 to 2007, and the longest period of 15 years to 2017. This illustrates the nonlinear and quasi-linear nature of the various profiles. In each case, these profiles could contain concave, linear or convex segments, reflecting the degree of re-ranking being experienced as h is increased to include higher-income individuals. A greater amount of re-ranking mobility tends to generate profiles that are more concave. That is, like the TIM curve, greater concavity implies more-equalising positional mobility. Convexity implies disequalising re-ranking, with neutrality captured by linear segments.

It can be seen in Figure 8 that the three re-ranking curves (absolute, positive and net) have similar shapes in both periods, but differ largely in the magnitudes of re-ranking as shown by the vertical axis scales. Note that the maximum re-ranking possible is 1 (positive and net) or 2 (absolute). Thus, for the whole population of individuals, absolute re-ranking reaches around 0.6 after 5 years and exceeds 0.8 after 15 years (at the 100^{th} percentile). Similarly positive re-ranking reaches around 0.3 and 0.4 respectively.

To assess the incidence, intensity and interpersonal aspects of these re-ranking measures, Figure 8 should be interpreted as follows. For a given definition of positional mobility (net, positive or absolute re-ranking), select a value of $h = k/n$ representing the subset of low-income individuals of interest (the incidence dimension). The height of the profile on the vertical axis at this value of h represents the intensity of re-ranking for this group; namely how much re-ranking they have experienced on average (or cumulatively). The section of the profile to the right of h becomes irrelevant.

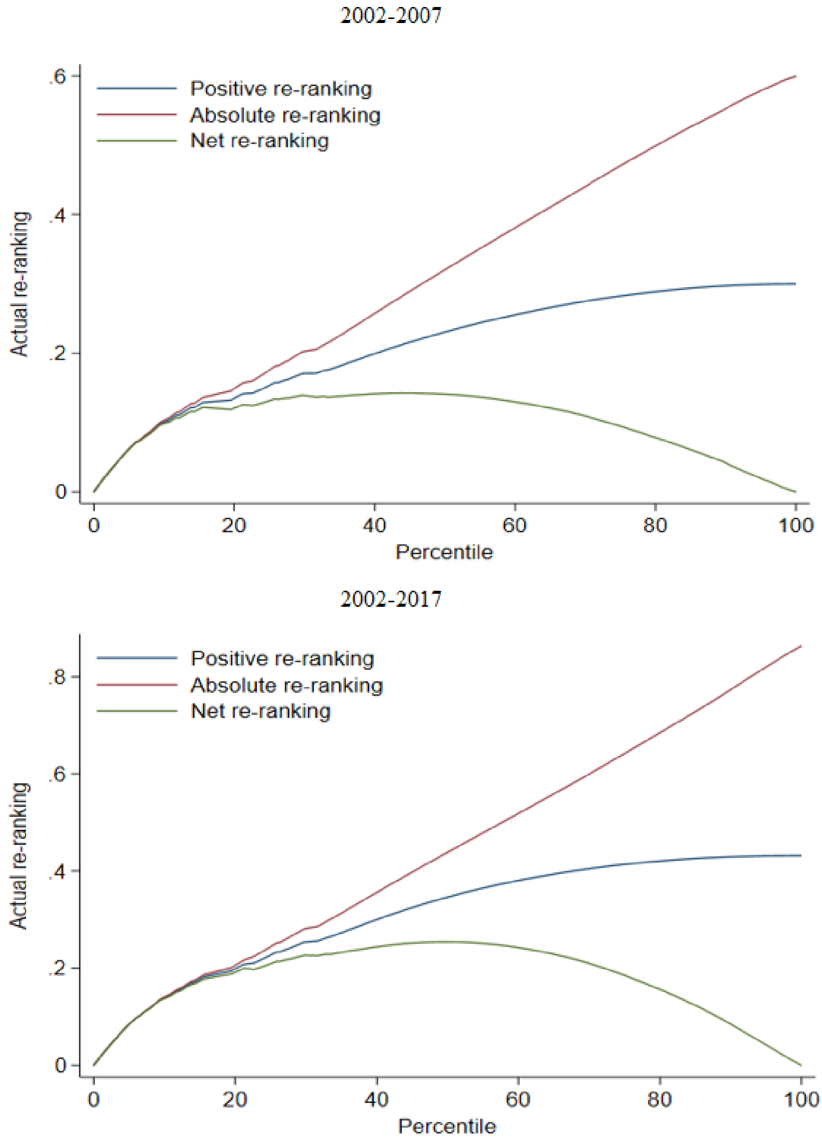

Figure 8: Re-rankng Curves, 2002-07 and 2002-17

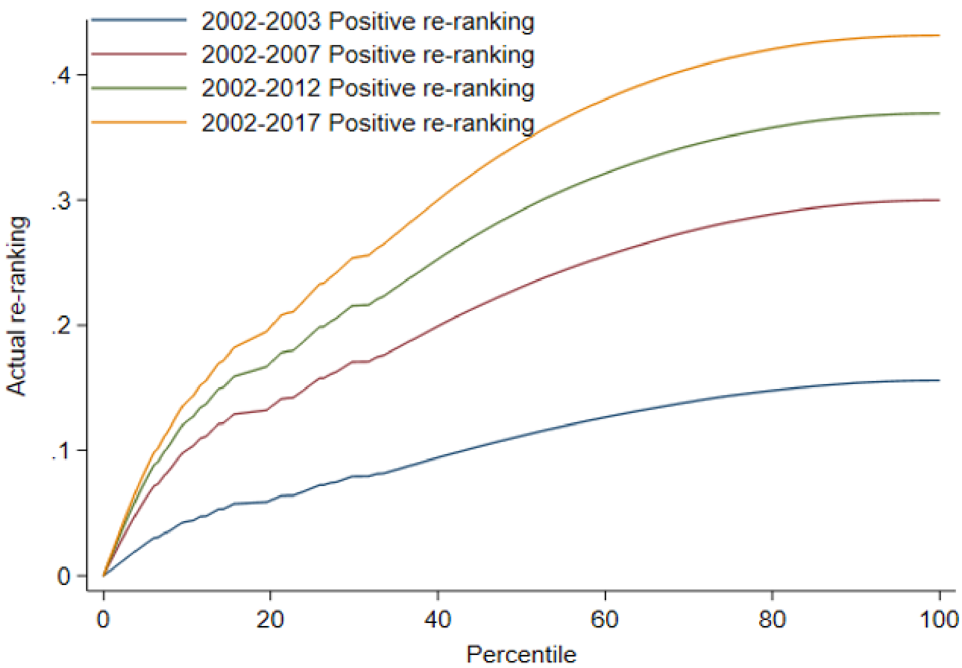

Figure 9: Absolute Re-ranking Across the Four Periods

The deviation from linearity of the m_k profile, from the origin to its value at the selected h, provides a measure of the degree of progressive (concave) or regressive (convex) re-ranking among individuals within the h^{th} percentile. That is, the actual profile may be compared to a straight line from the origin to the value of m_k at $h = 1$. For example, in Figure 8 the profiles for absolute re-ranking are remarkably linear, at least above the 10^{th} percentile. This suggests that, at least for this sample and measure, the extent of re-ranking is relatively constant across the income distribution.

As with the TIM curves in section 6, changes in the incidence, intensity and inequality of positional mobility associated with different time periods can be examined by plotting relevant m_k profiles for the four periods. Figure 9 illustrates this for the positive re-ranking measure, m_h^{pos}; absolute and net re-ranking measures display similar properties. As expected, these profiles shift upwards (indicating more re-ranking) the longer the period of time considered. The largest increase appears to be between the 1-year and 5-year periods, with total re-ranking at the 100^{th} percentile around 0.15 (15 per cent) after 1 year and 0.3 after 5 years. By 15 years this has reached over 0.4.

It is clear from Figure 9 that the characteristics of re-ranking mobility across the four periods are very similar in terms of the interpersonal dispersion of mobility (concavity) of each profile for any given percentile, h. Also, since the maximum positive re-ranking for $h \geqslant 0.5$ is equal to one (see Figure 2), the values of m_k in Figure 9 also reveal the values of the re-ranking ratio, $RRR_k = m_k/m_k(\max)$ for $h \geqslant 0.5$. The RRR_k profiles look very different at lower values of h, and for different net/positive/absolute concepts, as shown below. Thus, at $h = 1$, the value of m_h^{pos} in excess of 0.4 for 2002-2017 suggests that over the 15 years, more than 40 per cent of the maximum potential re-ranking occurred.

As shown above, while some groups of individuals in Figures 8 and 9 may experience higher re-ranking than others, their movements are constrained to differing degrees by the maximum re-ranking possible. However, the differences between the actual m_k and the equivalent $m_k(\max)$ can be identified by considering changes in RRR_k as $h \longrightarrow 1$. Re-ranking ratio curves, obtained by plotting RRR_k against h, are shown in Figure 10 for the three re-ranking measures over 2002-2007 (upper panel) and 2002-2017 (lower panel): values on the vertical axis are simply the ratios of the axis values in Figures 8 and 2.

The chart indicates that, for all three re-ranking measures in the New Zealand case, the extent of mobility relative to the maximum achievable is relatively high for the lowest-income individuals (low h), with $RRR_k \approx 0.4$ after 15 years for $h \approx 0.1$. This steadily declines as h is increased, reaching a minimum of approximately 0.3 at around the 20^{th} to 30^{th} percentiles, except in the case of the m_h^{net} profile which continues to decline but remains fairly flat for $h > 0.3$. Thereafter, the RRR_k^{abs} rises to around the 70^{th} percentile, while the RRR_k^{pos} profile continues to rise to the 100^{th} percentile.[26]

It may therefore be inferred that the group experiencing absolute re-ranking that is closest to the maximum achievable are the very low-income group and also the middle-income group between approximately the 40^{th} and 70^{th} percentiles where the RRR_k^{pos} curve is rising most steeply towards a (local) maximum at $h = 1$. For the positive re-ranking measure, the ratio of actual to maximum re-ranking is generally highest for both the low and high population percentiles, reaching around $RRR_k^{pos} = 0.3$ or more after 5 years and $RRR_k^{pos} = 0.40$ or more after 15 years. From Figure 10, the RRR_k^{pos} and the RRR_k^{abs} profiles reach the same value for $h = 1$. As Creedy and Gemmell (2019) show, this is not a coincidence, but reflects the properties of the two measures.

Considering the three profiles in Figure 10 it is clear that the measure of net movement, RRR_k^{net}, indicates a persistent downward trend as h moves towards 1. This suggests that the lowest-income individuals generally experienced more movement in their income rank over this period, relative to the maximum achievable, than those on higher incomes. This seems likely to be capturing a re-ranking analogue of the progressivity in income growth observed above.

[26]The strong fluctuations in the curves, at h close to 1, relfect the fact that the value of both the actual and maximum net re-ranking measures equal zero at $h = 1$. Hence the ratio can be quite unstable in the vacinity of $h = 1$ (and is undefined at $h = 1$).

The tendency for the ratio of actual to maximum possible re-ranking to rise, the longer the time period considered, can be seen for the positive re-ranking measure in Figure 11, which includes all four RRR_k^{pos} profiles. This reveals the volatility in RRR_k over the lowest 5 percentiles, perhaps not surprisingly given the numbers of individuals in each period on very low incomes in the initial year (for example, in 2002, the 5^{th} percentile income level is only around \$7,098), but who experience a wide range of income changes over the period. Much of this probably reflects some low-income individuals such as secondary earners, moving into employment or from part-time to full-time work, while others remain in their initial employment status. These data also include the self-employed who are known to experience greater annual income volatility.

All four profiles in Figure 11 behave similarly to the m_k^{pos} profiles in Figure 10, confirming greater re-ranking as a fraction of the maximum possible as more years are added. For example, the minimum RRR_k^{pos} occurs at around the 30^{th} percentile in all four profiles; it is approximately 0.1 after 1 year, rising to 0.3 after 15 years. Similarly the maximum RRR_k^{pos} values at very low percentiles rise from around 0.13 after 1 year to 0.45 after 15 years.

Figure 11 also suggests that the differences in RRR_k^{pos} across the percentiles tends to become more pronounced the longer the period considered. For example, re-ranking that occurs over just 1 year appears quite similar across all h values, at around 0.10 to 0.15. After 5 years, however, substantial differences across h values appear, ranging from 0.20 to 0.35. The results also demonstrate that, across an extended period of 15 years, positive re-ranking is typically around 30-45 percent of the maximum mobility possible, conditional on an individual's position in the initial income distribution. In all periods, it also tends to be highest at both the bottom and at the top of those initial distributions.

7.1 Re-ranking Ratio Curves by Gender, Ethnicity and Education

As with the TIM curves in Sections 6.2 and 6.3, changes in the incidence, intensity and inequality of positional mobility associated with different decompositions can be examined by plotting relevant re-ranking profiles over the short- and long-run. Figure 12 compares profiles for RRR^{pos} for males and females over the four periods. As can be seen, some volatility observes in RRR over the lowest 40^{th} percentiles for females. For males, on the other hand, such volatility only exists for the lowest 25 th percentiles. This perhaps is not surprising given that females are more likely to shift between full- and part-time jobs or moving in and out of employment for child care reasons. Note that across all four periods, females generally experience greater re-ranking as a fraction of the maximum possible compared to their males counterparts. The results also demonstrate that, over the shorter period, re-ranking is typically around 10 - 30% of the maximum mobility possible (conditional on an individual's position in the initial income distribution). As the period length is extended, this rate becomes closer to 50%.

Figure 13 demonstrates RRR^{pos}profiles for three different ethnic groups. This Figure suggests that re-ranking over a shorter period (e.g., 1 year) appears to be approximately linear across all h vlaues. This is more obvious for Māori ethnic group. From Figure 13, it is also apparent that the properties of re-ranking mobility across all three ethnic decompositions are similar in terms of the positional re-ranking mobility as the period considered is extended from 1 year to 15 years. However, the main difference between the three groups is that non-Māori and non-Pasifika groups experience the less amount of re-ranking mobility compared to the other two counterparts. The difference in re-ranking mobility decreases after 15 years.

Across-educational qualifications differences using RRR^{pos}, for the same periods, are shown in Figure 14. The Figure suggests that differences in re-ranking occured within the two ends of the spectrum, with no qualification and with University degree, appears to be similar across all h values. By 15 years, the two re-ranking curves for no qualification and university degree are almost indistinguishable for those above 70^{th} percentile. Note that the minimum of RRR^{pos}occurs at around the 20^{th} percentile in all four periods for the school- and post-school qualifications. This happends at around the 30^{th} percentile for without qualification and university degree cases. In general, a greater amount of re-ranking mobility is observed with the higher qualification attained.

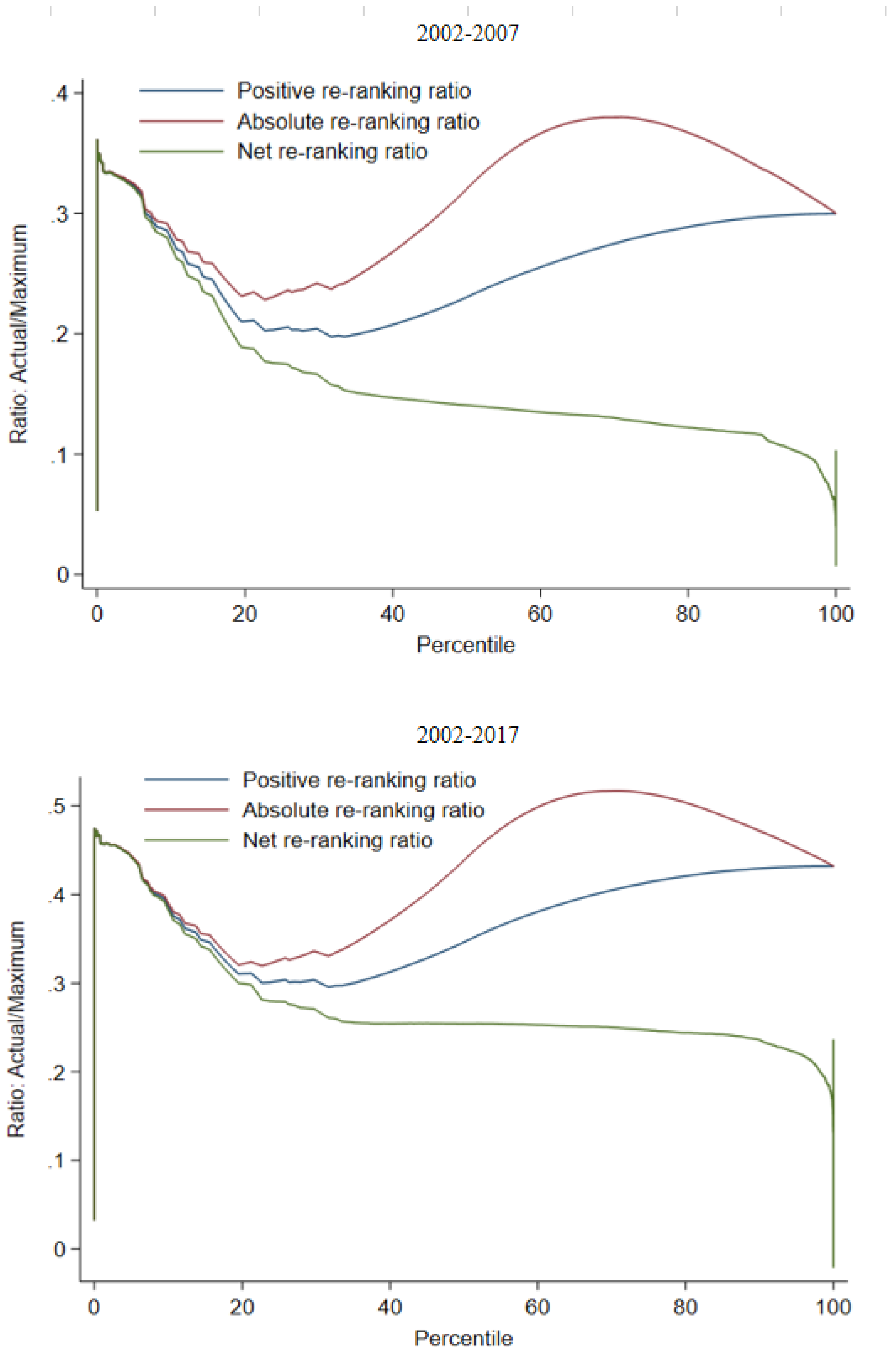

Figure 10: Re-ranking Ratio Curves, 2002-07 and 2002-17

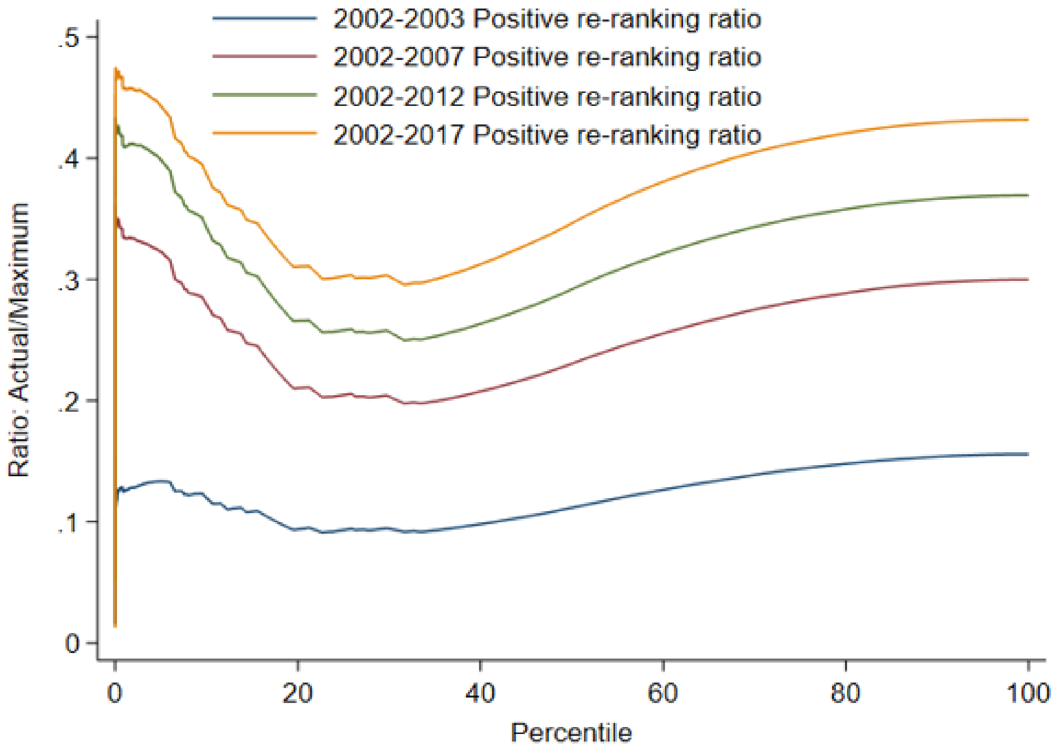

Figure 11: Positive Re-ranking Ratio Curves for Four Periods, 2002 to 2017

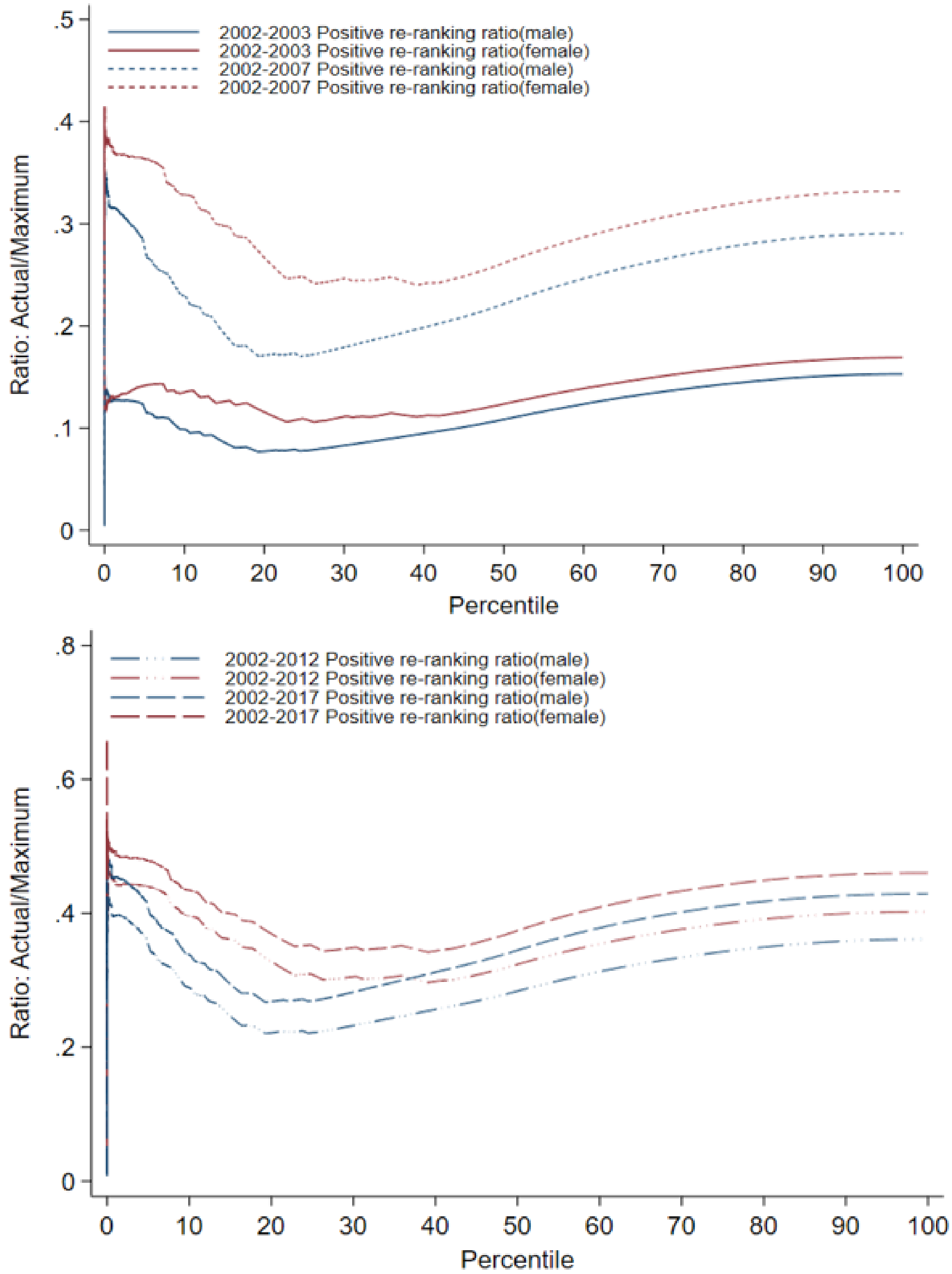

Figure 12: Comparing Positive Re-ranking Ratio Curves for Males and Females

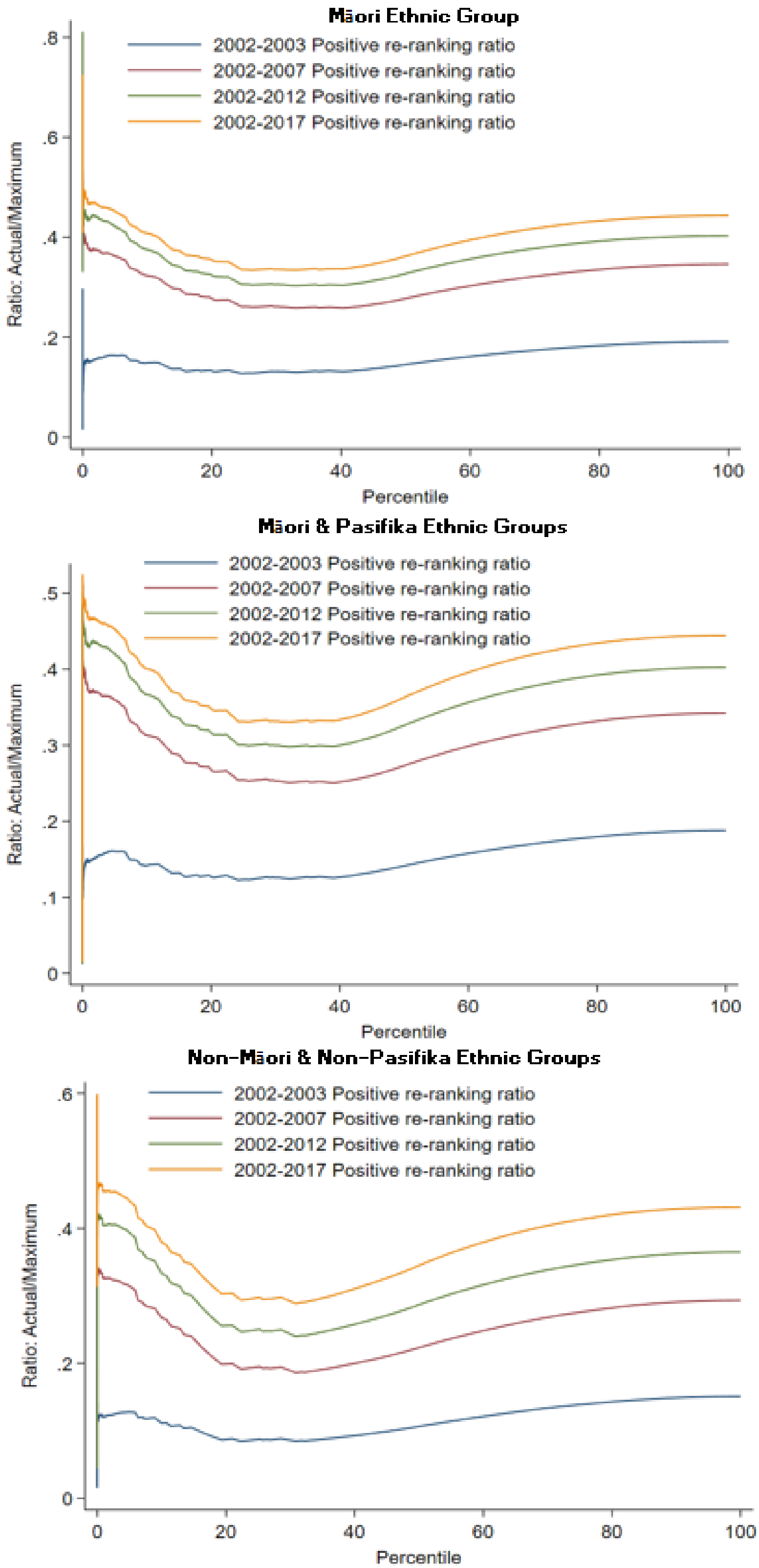

Figure 13: Comparing Positive Re-ranking Ratio Curves for Ethnic Groups

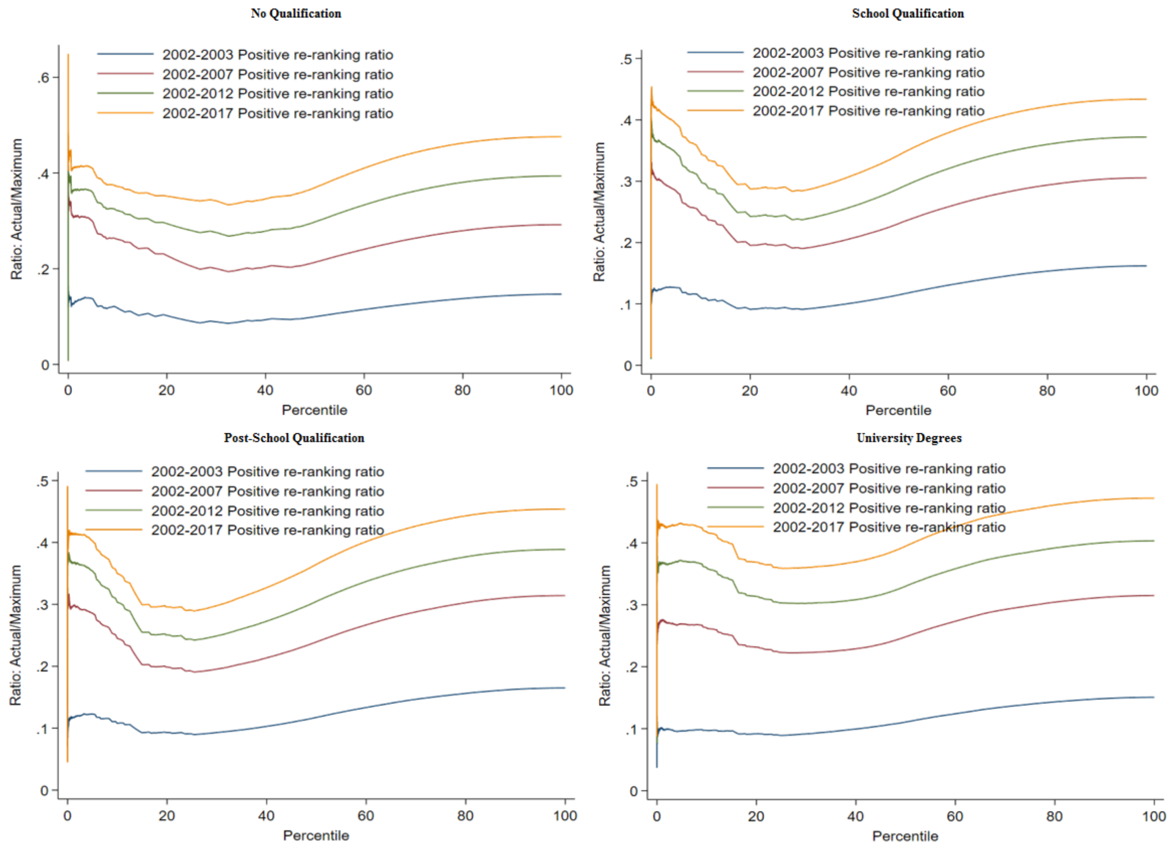

Figure 14: Comparing Positive Re-ranking Ratio Curves for Educational Qualifications

8 Conclusions

Relative income growth and positional change approaches to measuring income mobility have been applied here to extensive administrative longitudinal data on the taxable incomes of New Zealand taxpayers over a number of periods ranging from 1 year to 15 years. Illustrations for both mobility concepts were presented based on panel data for 2002-03, 2002-07, 2002-12 and 2002-17. These showed that income growth rates within the lower part of the income distribution were quite substantially higher than those observed higher up the income distribution; that is, income growth was 'progressive', reflecting in part a relatively high degree of 'regression towards the mean'.

After 15 years, evidence on the extent of re-ranking of individual incomes suggested a relatively high degree of positional mobility, compared to the maximum possible, especially among the lowest and highest income individuals. The evidence also suggested that some conclusions regarding the extent of re-ranking depends crucially on the re-ranking measure adopted – positive, net or absolute. For example, the highest re-ranking ratios are observed around the 40^{th} to the 70^{th} percentiles for an absolute re-ranking measure but rise steadily towards the 100^{th} percentile when only positive changes in rank are considered.

Finally, the evidence here for New Zealand, that income growth from this longitudinal perspective is progressive, is consistent with that found by Van Kerm (2009) and Jenkins and Van Kerm (2011, 2016) using their income growth profiles for the UK and a selection of other European countries. That is, in each case income growth rates are generally greater for those initially on lower incomes. Jenkins and Van Kerm (2016) note that this stands in contrast to evidence from repeated cross-sections, identified using UK growth incidence curves for example. Though the present paper has not examined cross-sectional data on income growth, Creedy and Gemmell (2019b) found that this cross-sectional/longitudinal difference also appears to hold for New Zealand.

References

[1] Alinaghi, N., Creedy, J. and Gemmell, N. (2020) Constructing a longitudinal database for the analysis of individual incomes in New Zealand. *Working Papers in Public Finance*, 05/2020. Wellington School of Business and Government, Victoria University of Wellington, Wellington, New Zealand.

[2] Alinaghi, N., Creedy, J. and Gemmell, N. (2022a) Income inequality and the accounting period in New Zealand: evidence from administrative Data. *Working Papers in Public Finance* No. 05/2022, Victoria School of Business and Government, Victoria University of Wellington, Wellington, New Zealand.

[3] Alinaghi, N., Creedy, J. and Gemmell, N. (2022b) Inter-decile movements of individuals in New Zealand: evidence from administrative data. *Working Papers in Public Finance* No. 4/2022, Victoria School of Business and Government, Victoria University of Wellington, Wellington, New Zealand.

[4] Atkinson, A.B. and Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status. *Review of Economic Studies*, 49, 183-201.

[5] Bourguignon, F. (2011) Non-anonymous growth incidence curves, income mobility and social welfare dominance. *Journal of Income Inequality*, 9, 605-627.

[6] Cowell, F.A. and Flachaire, E. (2018) Measuring mobility. *Quantitative Economics*, 9, 865-901.

[7] Creedy, J. and Gemmell, N. (2017) Illustrating income mobility: two new measures. *Working Papers in Public Finance* No. 12/2017, Victoria School of Business and Government, Victoria University of Wellington, Wellington, New Zealand.

[8] Creedy, J. and Gemmell, N. (2018) Income dynamics, pro-poor mobility and poverty persistence curves. *Economic Record*, 94, 306, 316-328.

[9] Creedy, J. and Gemmell, N. (2019a) Illustrating income mobility: new measures. *Oxford Economic Papers*, 71, 733-755.

[10] Creedy, J. and Gemmell, N. (2019b) Income inequality in New Zealand: why conventional estimates are misleading. *Agenda—A Journal of Policy Analysis and Reform*, 26, 5-22.

[11] Creedy, J., Gemmell, N. and Laws, A. (2021) Relative income dynamics of individuals in New Zealand. *New Zealand Economic Papers*, 55, 203-220.

[12] D'Agostino, M. and Dardanoni, V. (2009). The measurement of rank mobility. *Journal of economic theory*, 144(4), 1783-1803.

[13] Fields, G.S. (2000). *Income mobility: Concepts and measures: Patterns and underlying causes*. Cornell University, ILR School.

[14] Grimm, M. (2007) Removing the anonymity axion in assessing pro-poor growth. *Journal of Economic Inequality*, 5, 179-197.

[15] Jäntti, M. and Jenkins, S.P. (2015) Income Mobility. In *Handbook of Income Distribution*, Volume 2 (edited by A.B. Atkinson and F. Bourguignon. New York: Elsevier).

[16] Jenkins, S.P. and Van Kerm, P. (2011) Trends in individual income growth: measurement methods and British evidence. *IZA Discussion Paper* DP No. 5510.

[17] Jenkins, S.P. and Van Kerm, P. (2016) Assessing individual income growth. *Economica*, 83, 679-703.

[18] Jenkins, S.P. and Lambert, P.J. (1997) Three "I"s of poverty curves, with an analysis of UK poverty trends. *Oxford Economic Papers*, 49, 317-27.

[19] Markandya, A. (1982) Intergenerational exchange mobility and economic welfare. *European Economic Review*, 17, 307-324.

[20] Palmisano, F. and Peragine, V. (2015) The distributional incidence of growth: a social welfare approach. *Review of Income and Wealth*, 61, 440-464.

[21] Ravallion, M. and Chen, S. (2003) Measuring pro-poor growth. *Economics Letters*, 78, pp. 93-99.

[22] Shorrocks, A.F. (1978) Income inequality and income mobility. *Journal of Economic Theory*, 19, 376-393.

[23] Simandan, D. (2018) Rethinking the health consequences of social class and social mobility. *Social Science & Medicine*, 200, 258-261.

[24] Son, H.H. (2004) A note on pro-poor growth. *Economics Letters*, 82, 307-314.

[25] Son, H.H. (2007) *Pro-Poor Growth: Concepts and Measures.* Asian Development Bank, Economics and Research Department. ERD Techncial Note Series, No. 22.

[26] Svallfors, S. (ed.) (2005) *Life Changes and Social Mobility in Comparative Perspective.* Stanford, CA: stanford University Press.

[27] Trede, M. (1998) Making mobility visible: a graphical device. *Economics Letters*, 59, pp. 77-82.

[28] Van Kerm, P. (2009) Income mobility profiles. *Economics Letters*, 102, 93-95.