# Age-Income Profiles in New Zealand: New Estimates Based on Administrative Data

Nazila Alinaghi, John Creedy and Norman Gemmell

WORKING PAPER 08/2022
March 2022

Working Papers in Public Finance

VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

Chair in Public Finance
Wellington School of Business and Government

**The Working Papers in Public Finance** series is published by the Victoria Business School to disseminate initial research on public finance topics, from economists, accountants, finance, law and tax specialists, to a wider audience. Any opinions and views expressed in these papers are those of the author(s). They should not be attributed to Victoria University of Wellington or the sponsors of the Chair in Public Finance.

Further enquiries to:
The Administrator
Chair in Public Finance
Victoria University of Wellington
PO Box 600
Wellington 6041
New Zealand

Phone: +64-4-463-9656
Email: cpf-info@vuw.ac.nz

Papers in the series can be downloaded from the following website:
https://www.wgtn.ac.nz/sacl/centres-and-chairs/cpf/publications/working-papers

# Age-Income Profiles in New Zealand: New Estimates based on Administrative Data*

Nazila Alinaghi, John Creedy and Norman Gemmell†

**Abstract**

This paper uses a new longitudinal dataset, containing information about the incomes of New Zealand individuals, to examine the form of cohort age-income profiles. A model of the variation in mean log-earnings with age, allowing for quadratic age and linear time effect, is estimated separately for males and females, along with a range of other demographic groups. An 'overtaking' property, whereby more recent cohorts have higher real income than older cohorts, at comparable ages, are found in all cases. Cubic profiles of the variation in the variance of log-income with age are also estimated. Examples of the projected changing distribution of income with age are given, for various cohorts aged 20 in 2020.

**JEL Classification:**

**Keywords:** Taxable income; age-income profiles; cohort incomes

*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 here are based in part on tax data supplied, for statistical purposes, to Statisticss New Zealand (Stats NZ) by the Inland Revenue Department, under the Tax Administration Act 1994. Any discussion of data limitations or weakness is in the context of using the Integrated Data Infrastructure 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 Stats NZ 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 Stats NZ 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 Stats NZ. More information about these databases can be obtained at: *https://www.stats.govt.nz/integrated-data/*.

# 1 Introduction

It has long been recognised that there are systematic variations over the life cycle in the income distribution of different cohorts, defined by year of birth. These variations significantly affect, among other things, consumption and savings behaviour, and complicate the interpretation of aggregate measures of income inequality. In addition, the life-cycle income patterns are known to differ according to gender, education and ethnicity. This means that changes in the inequality of income, measured over all individuals, can change over time simply as a result of changes in the age composition of the population.

Early attempts to measure life-cycle income varations in New Zealand were severely contrained by data availability. For many years, even when published data could be disaggregated by age groups, only cross-sectional comparisons were possible. Such cross-sectional profiles, which essentially compare incomes in different cohorts, do not necessarily reflect the experience of separate cohorts: they compare incomes of people of different ages whose incomes are measured at a single calendar date. Earlier cohort analyses, using specially-constructed anonymised datasets based on Inland Revenue data, were restricted by the relatively short length of time over which incomes were traced, and the ability to distinguish only between males and female cohorts. This was because tax administration does not require other demographic information from individuals.[1]

The present paper examines age-income profiles of individuals in various demographic groups in New Zealand, using information from a special dataset compiled using the Statistics New Zealand Integrated Data Infrastructure (IDI). The present study makes use of longitudinal data on the annual taxable incomes of individuals, over the 18-year period from 2000 to 2017 inclusive. Hence, the experience of separate cohorts, over different parts of the life cycle, can be compared. By linking Inland Revenue data on taxable incomes with other administrative datasets, it is also possible to construct separate profiles for a range of demographic groups.

Any analysis of age-income profiles must recognise that, even where in-

[1]For an early analysis of cohort profiles in New Zealand, see Creedy (1997).

formation is available about the incomes of particular cohorts as they age over a range of calendar time, estimation of the variation that is associated purely with ageing requires very strong assumptions. This is because of the possible existence of separate age, cohort and calendar-time effects. People in different cohorts overlap at a particular calendar time, and changes in labour market conditions may affect those cohorts differently. Allowance for one particular influence – the rate of inflation – can be made by measuring all incomes in 'real' terms.

Summary information about the data used here is provided in Section 2. Section 3 provides diagrammatic illustrations of age-income profiles, and goes on to present and estimate a formal statistical model of the variation in mean log-income. Section 4 extends the analysis to deal with the variance of log-income, and reports projections of alternative age-income profiles for different cohorts. Brief conclusions are in Section 5.

# 2 The Data

The construction of the dataset of individual incomes is described in detail in Alinaghi *et al.* (2020) and is summarised only briefly here. The dataset has been made possible due to the improved availability of anonymised administrative register data, such as from individuals' tax records, in Statistics 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 reduced measurement error. Information about a range of demographic characteristics of individuals was obtained by linking Inland Revenue data on taxable incomes with several other administrative datasets contained within the IDI.

An obvious limitation of the dataset is that, because the income data are originally collected for tax administration purposes, the data cannot capture those who do not interact with the income tax system. Furthermore, there is necessarily an absence of information on non-taxable income. While recognising these limitations, the newly-constructed dataset used in this paper nevertheless provides the most comprehensive information to date on New Zealand taxpayers' incomes, suitable for the analysis of age-income profiles.

As mentioned above, the final dataset employed in this study was obtained by merging a number of administrative datasets within the IDI. The primary database covers the Inland Revenue individual taxpayer population, containing detailed tax return information about 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 5,393,874 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 1,447,755 individuals is available with income data for all 18 years. This forms the 'base' dataset.[2]

Ethnicity is a self-identified concept and individuals may specify multiple ethnicities. For present purposes, a 'prioritised ethnicity' variable is constructed. This means that individuals are classified into only one ethnic group in a prioritised order of Māori, Pasifika, Asian, European, MELAA, and Other. To construct this variable, an individual is classified as Māori, if their ethnic code in one of the data sources where this information is recorded is Māori. This process is repeated for other ethnic groups in order; for further details see Alinaghi *et al.* (2020, pp. 11-12).

Decompositions of this sample of the taxpayer population with annual data over the 2000 to 2017 are shown in Table 1, for gender, ethnicity, and highest educational qualification. This indicates that the gender composition is close to 50:50 between males and females; Māori and Pasifika represent 14 per cent and 5 per cent respectively of all individuals. 'Non-Māori, non-Pasifika' ethnicities recorded in the dataset include European, Asian, MELLA (Middle Eastern, Latin American, and African) and Other.[3]

[2]However, the regression analyses of age-income profiles reported here are restricted to individuals aged from 20 to 65 years (defined by age in 2000) yielding a dataset of 1,249,116 individuals.

[3]In the 2018 New Zealand Census, out of a total population of 4,699,755 individuals, ethnicity percentage were as follows: Māori (17), Pasifika (8), European (70), Asian (15),

Data on highest educational qualifications are constructed such that an individual is assigned to a category according to their highest qualification obtained in any year during the period of study, 2000 to 2017. For example, an individual obtaining a university degree in any given year, say 2013, is allocated to this category throughout the whole period. This avoids changes in sub-sample sizes for each qualification category during the period, and reflects an individual's educational capability or potential. Table 1 shows that there are 17 per cent of individuals with no qualification, close to those with 'University' degrees (16 per cent). Individuals with 'School' and 'Post-school' qualifications represent 32 and 23 per cent of the total respectively. 'Post-school' qualification includes diplomas and certificates awarded by educational and training institutions such as technical colleges. It also includes on-the-job training certificates.

Table 1: Sample Sizes by Decomposition

| | |
|---|---|
| Gender: | |
| Male | 645,960 |
| Female | 603,156 |
| Ethnicity: | |
| Māori | 178,677 |
| Pasifika | 58,572 |
| Non-Māori, non-Pasifika | 1,011,867 |
| Highest educational qualification*: | |
| None | 207,570 |
| School | 404,181 |
| Post-school | 287,049 |
| University | 196,290 |
| Total | 1,249,116 |

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

One characteristic that, unfortunately, is missing from the present analysis is the occupation of individuals. This variable in not available in the administrative data used to compile the special dataset used here. Further-

MELLA (1), and Others (1).

more, the occupational classification of individuals can vary over their life cycle.

# 3 Age-Income Profiles

The main features of age-income profiles, and differences between demographic groups, are first displayed diagrammatically in subsection 3.1. Subsection 3.2 goes on to present a formal model of cohort profiles. As mentioned above, this first requires a decision to be made about assumptions regarding the nature of the 'age', 'time' and 'cohort' effects on incomes. This subsection specifies a simple model, in which the profiles of different cohorts are assumed to differ only because those individuals age over different calendar time periods. Estimates are provided in subsection 3.3.

## 3.1 Age-Income Profiles Illustrated

Figure 1 shows, for all males and females, the variation in mean log-income with age, for a range of cohorts, defined by year of birth. To avoid too much clutter, the diagrams show profiles for cohorts separated by five years. These profiles show the characteristic 'humped' shape, with somewhat flatter and lower profiles for women. In contrast with some earlier data, the profiles for women do not appear to have a cubic form; see Creedy (1997), Creedy, Gemmell and Laws (2021). For relevant cohorts, the profiles initially shift upwards slightly after the traditional 'retirement age' (and age of eligibility for New Zealand Superannuation) of 65. This may reflect a tendency for retirees to transform previously accumulated assets into income soon after retirment and/or receipt of additional income in the form of 'retirement packages' from employers.

The profiles also reveal the phenomenon of 'overtaking', whereby younger cohorts have higher real incomes, at comparable ages, compared with older cohorts. However, for more recent male cohorts this is not pronounced. A larger extent of overtaking is shown for the older cohorts, particularly in 'post retirement' age groups: this probably reflects lower levels of wealth and hence non-wage incomes for the older male cohorts. Overtaking among female cohorts is evident across all cohorts.

Figure 2 shows age-income profiles for all males and females, excluding Maori and Pasifika. Not surprisingly, given the demographic composition of the population, these profiles are similar to those obtained for all males and females. Comparable profiles for Maori are illustrated in Figure 3. For males, the profiles are lower and somewhat flatter. For females, the profiles are lower than for 'all women', but the difference is less than for men.

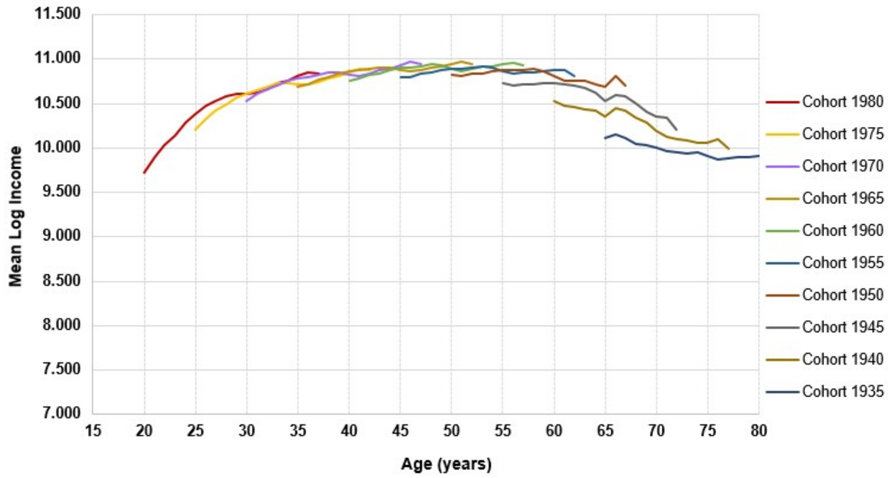
Age-income profiles for groups distinguished by highest educational qualification level are shown in Figures 4, 5 and 6.[4] For males, the profiles are lower and flatter than for 'all males', as expected, and except for the older cohorts they display less overtaking. However, this real-income overtaking is much more pronounced for women with only school-level qualifications. Again, the upward shift in profiles at age 65 is evident, as is the greater overtaking among the older cohorts. For those with a university degree, the profiles are somewhat more 'peaked' than for the other groups. Again this is an expected feature of age-income profiles for more highly educated groups.

## 3.2 The Variation in Average log-Income with Age

The age-income profiles discussed in the previous subsection suggest that, in logarithmic form, they can be approximated by a quadratic function. However, in specifying the precise form, a decision must first be made regarding 'cohort' and 'age' effects. The assumption made here is that there are no cohort effects. This means that there are assumed to be no distinctive features of particular cohorts (such as their size) that have independent effects on the systematic growth of members' incomes. As shown below, this does *not* mean that the age-income profiles for different cohorts – which age over different time periods – are the same. The assumption about cohort effects is made because comparisons among cohorts generally show considerable stability in

[4] An individual is included in the relevant highest educational qualification category throughout the period under consideration, regardless of the year in which that qualification is obtained. For example, an individual obtaining a university degree in 2006 is asigned to this educational category in all 4 transition periods over 2002 to 2017.

Males




Females

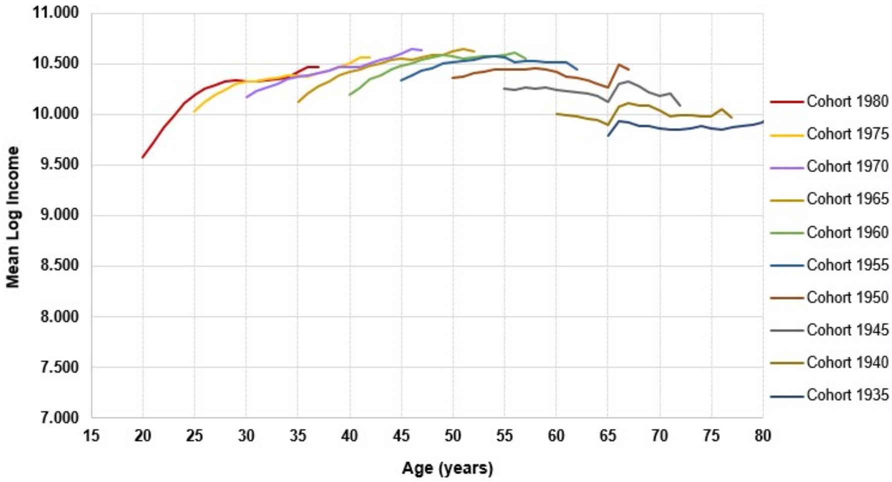



Figure 1: Cohort Income Profiles: All Males and Females

Males

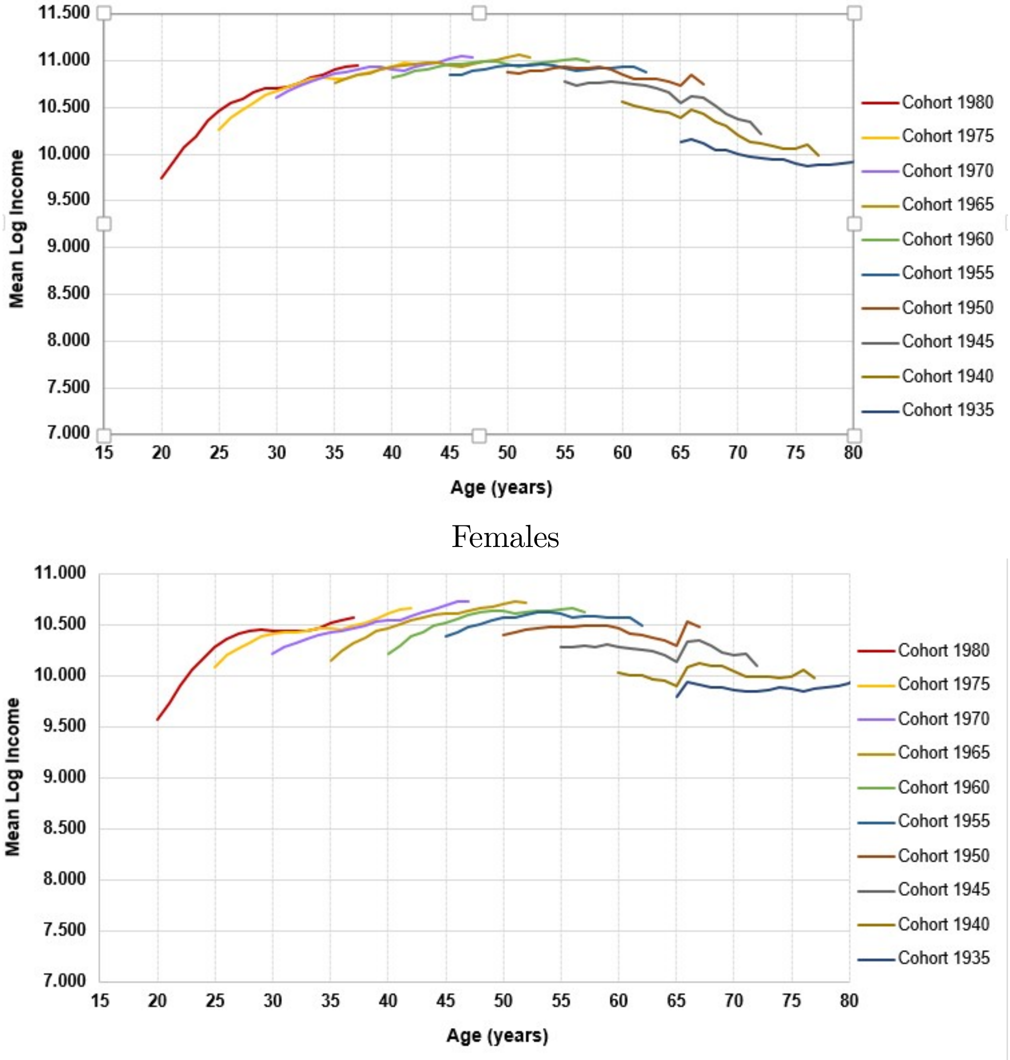



Figure 2: Cohort Income Profiles: All Excluding Maori and Pasifika Males and Females

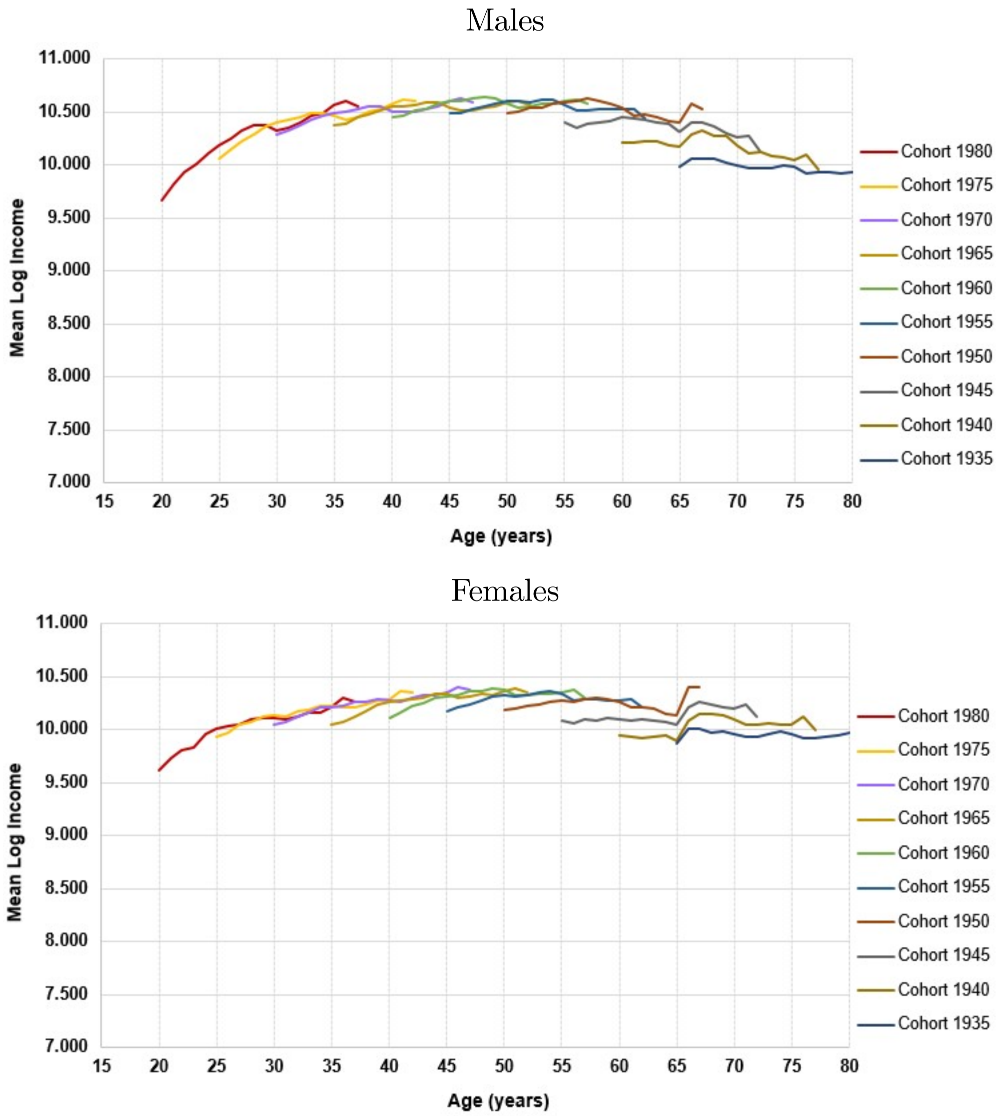



Figure 3: Cohort Income Profiles: Maori Males and Females

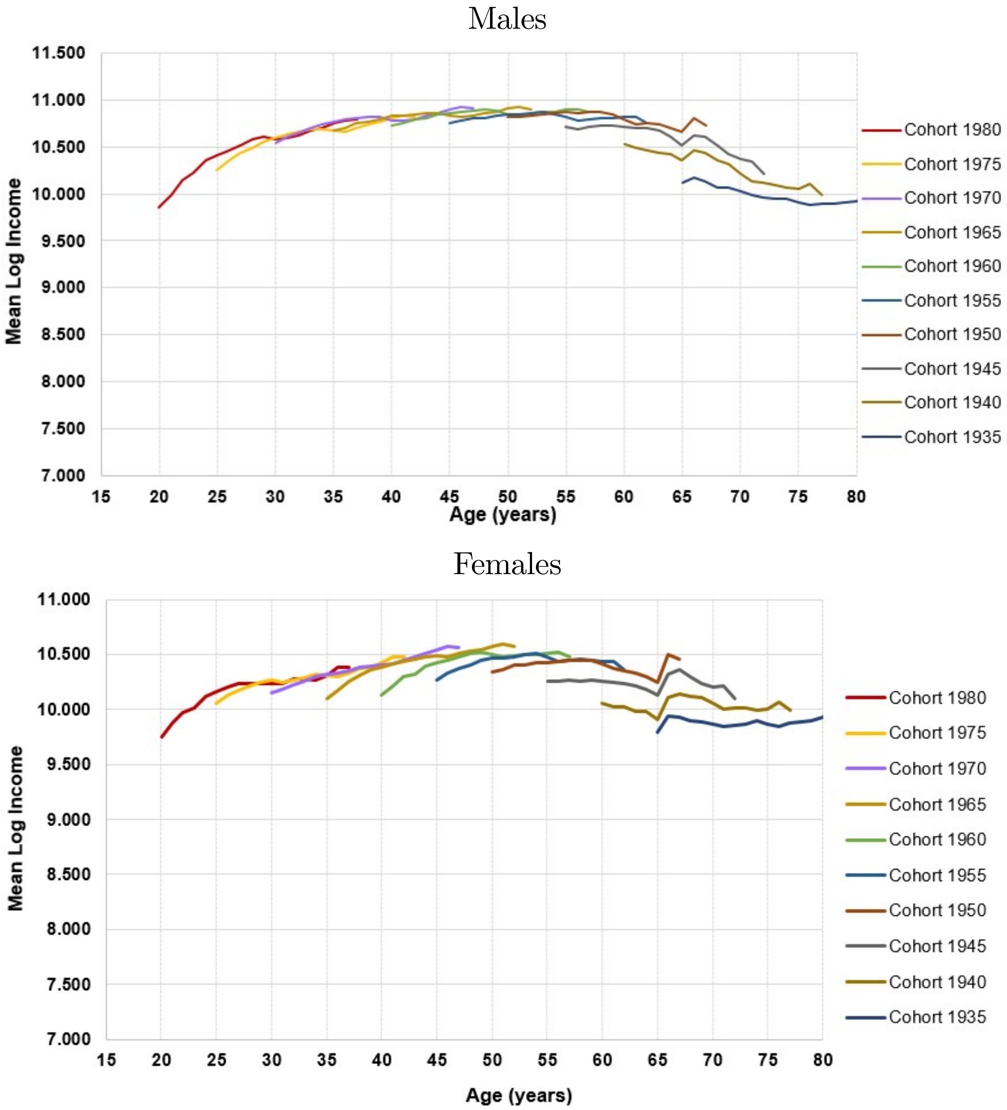
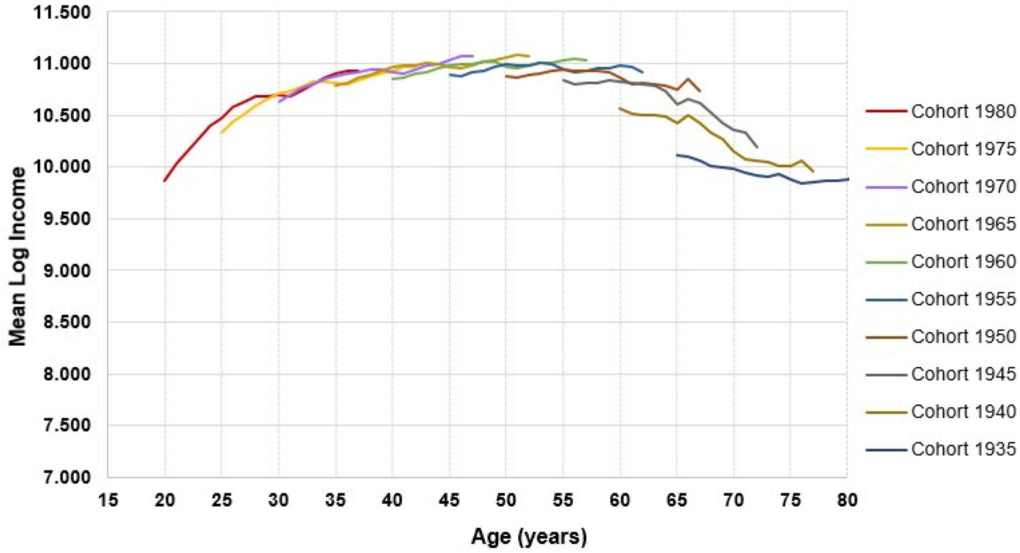



Figure 4: Cohort Income Profiles: School Qualification Only

Males




Females

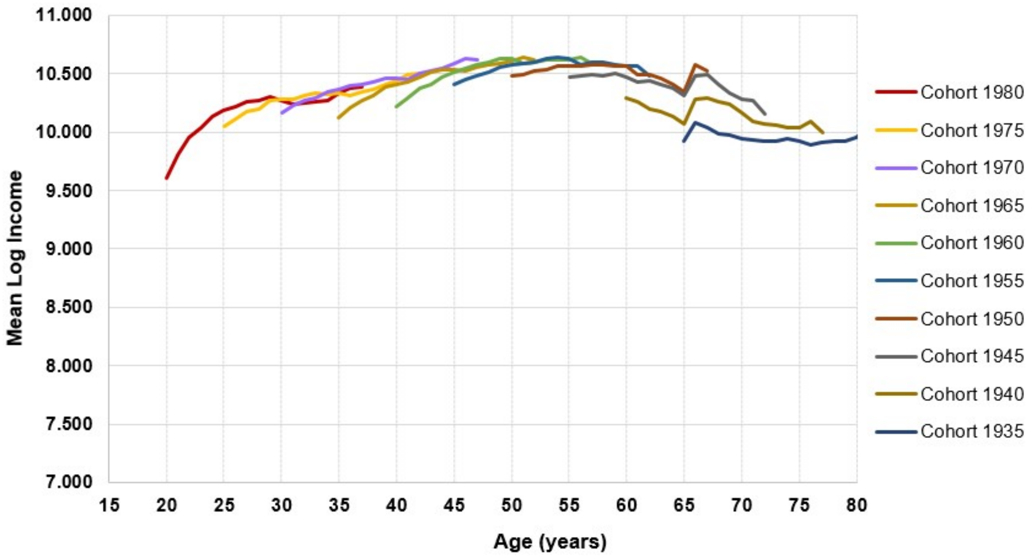
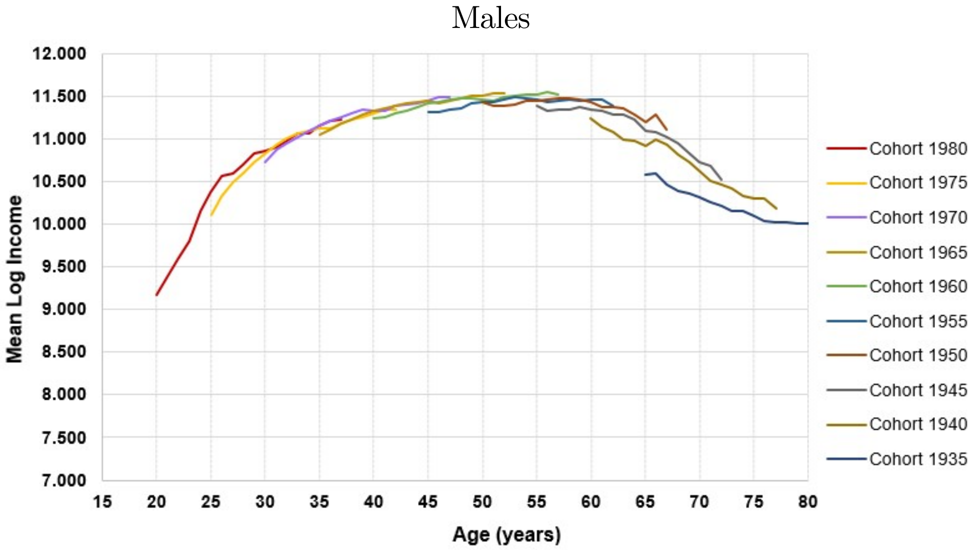



Figure 5: Cohort Income Profiles: Post School Qualification



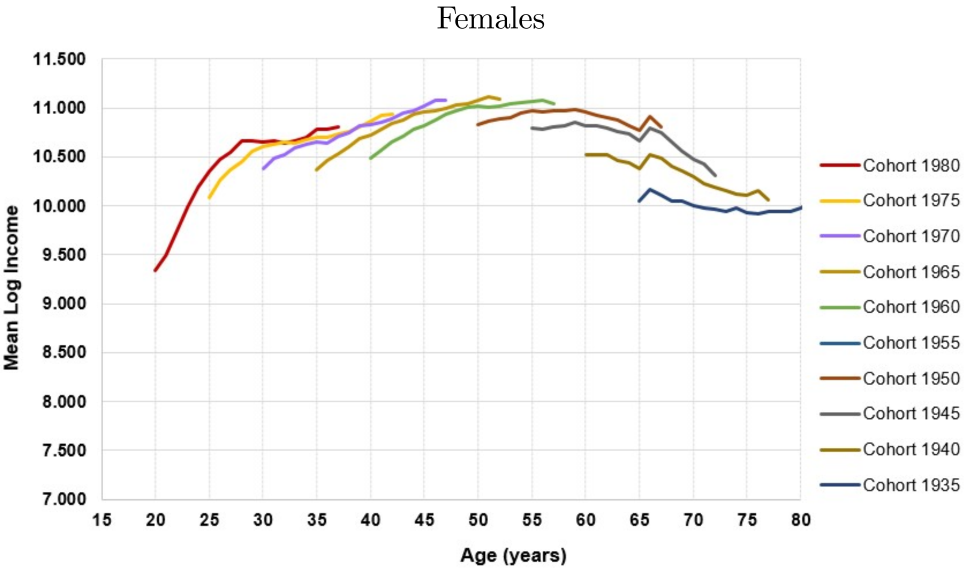




Figure 6: Cohort Income Profiles: University Degree

the typical form of their age-income profiles, and few cohort characteristics are readily available.[5]

The second set of assumptions refer to the specification of the 'age' and 'calendar time' effects on income growth. Following extensive previous evidence, and based on the previous subsection, the age effects are assumed to be quadratic.[6] The specification here is applied only to ages below 65, in view of the complications revealed above for years following the traditional retirement age. Following earlier results, the calendar time effects are assumed to be linear.[7] The specification in terms of the mean of log-income implies that arithmetic mean income follows a more complex profile which depends on the nature of the income distribution within each age group.[8]

With these assumptions, the mean of log-income at age, $t$, and calendar date, $s$, $\mu_{t,s}$, can be specified as:[9]

$$\mu_{t,s} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta s \tag{1}$$

The parameter, $\alpha_2$, is expected to be negative, consistent with the familiar concave shape of earnings profiles. The way in which the arithmetic mean log-income of a particular cohort varies with the age of the cohort can be seen by using the fact that $s - t$ gives the date of birth of a cohort, $c$. For example, people aged 20 in 2015 were born in 1995. Hence, $s = t + c$, and substitution into 1 gives the variation in $\mu_t$ with age, $t$, for a particular cohort as:

$$\mu_{t|_c} = (\alpha_0 + \beta c) + (\alpha_1 + \beta)\, t + \alpha_2 t^2 \tag{2}$$

This results shows that both the intercept and the coefficient on age can differ between cohorts, depending on the time effects, even though no specific cohort effects are assumed to operate. The relevant parameters are estimated

[5] However, some early studies found cohort effects in the US; see, for example, Weiss and Lillard (1978), Welch (1979), Connolly (1986) and Berger (1985).

[6] It is possible to relate the quadratic profile to a simple model of the returns to education; see Creedy (1985).

[7] Creedy (1997) found that quadratic terms for calendar time were very small and not significantly different from zero.

[8] This is discussed further below, in the context of the two-parameter lognormal distribution.

[9] This follows Creedy (1985, p. 76; 1992, pp. 81-90).

directly by carrying out a regression using equation (1). The positive time effect, $\beta$, produces the overtaking phenomenon observed earlier, whereby the earnings profiles of younger cohorts are higher, at comparable ages, than those of older cohorts.

### 3.3 Empirical Estimates

This subsection reports estimates of the specification in equation (1). Following the same procedure as in Creedy (1997), regressions are carried out for different demographic groups (males and females seperately), using the observed values of $\mu$ for 23 cohorts aged 20-21 up to 64-65 in 2000. The values of $t$ are adjusted so that $t = 0$ for age 20 and therefore, $t$ is allowed to vary from 0.5 to 61.5. The values of $s$ (calendar dates) are adjusted so that they take the values 1, 2, ... 18 for 2000, 2001 ... 2017, respectively. Hence, for the first cohort (those aged 20-21 in 2000) eighteen values of $\mu$ are available corresponding to $(t, s)$ combinations of (0.5, 1), (1.5, 2), ..., and (17.5, 18), respectively. Similarly, for the second cohort (those aged 22-23 in 2000), the eighteen observed values of $\mu$ corresponds to $(t, s)$ combinations of (2.5, 1), (3.5, 2), ..., and (19.5, 18), respectively. As a result, each cohort has sixteen values of age, $t$, in common with its immediately younger cohort. Since those values corresponds to a different set of calendar time values, $s$, cohort values of $\mu$ are not necessarily expected to be the same (unless time effects are zero). Thus, for each regression, there are 414 observations (23 cohorts for 18 years). The values of $\mu$ used in the regressions are the observed values after adjusting for inflation.

Table 2: Age-Income Profiles: Males and Females

| Coefficient | Estimate | t-value | Estimate | t-value |
|---|---|---|---|---|
| | Male: All | | Female: All | |
| C | 10.0168 | 545.65 | 9.8467 | 557.36 |
| Age | 0.0592 | 47.37 | 0.0362 | 30.10 |
| $Age^2$ | -0.0011 | -58.84 | -0.0007 | -39.23 |
| Date | 0.0141 | 16.58 | 0.0224 | 27.40 |
| $R^2$ | 0.9240 | | 0.8741 | |
| | Male Maori | | Female Maori | |
| C | 9.9694 | 704.38 | 9.8238 | 722.69 |
| Age | 0.0383 | 39.78 | 0.0246 | 26.61 |
| $Age^2$ | -0.0007 | -48.5 | -0.0005 | -32.44 |
| Date | 0.0108 | 16.42 | 0.0156 | 24.87 |
| $R^2$ | 0.8876 | | 0.8141 | |
| | Male Non-M + Non-P | | Female Non-M + Non-P | |
| C | 9.9912 | 586.6 | 9.8953 | 666.22 |
| Age | 0.0509 | 43.91 | 0.0274 | 27.07 |
| $Age^2$ | -0.0011 | -59.1 | -0.0007 | -43.37 |
| Date | 0.0272 | 34.52 | 0.0360 | 52.42 |
| $R^2$ | 0.7632 | | 0.7614 | |
| | Male School | | Female School | |
| C | 10.0617 | 549.41 | 9.8452 | 593.56 |
| Age | 0.0551 | 44.17 | 0.0331 | 29.31 |
| $Age^2$ | -0.0011 | -54.78 | -0.0007 | -37.05 |
| Date | 0.0116 | 13.63 | 0.0184 | 23.96 |
| $R^2$ | 0.9129 | | 0.8497 | |
| | Male Post School | | Female Post School | |
| C | 10.0905 | 493.52 | 9.7773 | 533.80 |
| Age | 0.0619 | 44.43 | 0.0461 | 36.95 |
| $Age^2$ | -0.0012 | -56.67 | -0.0008 | -42.24 |
| Date | 0.0149 | 15.73 | 0.0134 | 15.81 |
| $R^2$ | 0.9242 | | 0.8402 | |
| | Male Uni Degree | | Female Uni Degree | |
| C | 9.7840 | 416.3 | 9.7309 | 417.01 |
| Age | 0.1103 | 68.88 | 0.0727 | 45.75 |
| $Age^2$ | -0.0019 | -77.14 | -0.0013 | -52.85 |
| Date | 0.0146 | 13.4 | 0.0230 | 21.27 |
| $R^2$ | 0.941 | | 0.8946 | |

Table 2 gives results for the demographic groups examined above. In all cases the goodness of fit, as measured by $R^2$, is very high, and all coefficients are highly significantly different from zero, as shown by the high $t$-values. These results, together with similar regression results for log-income variances (see subsection 4.1), allow age-income profiles for individuals across the income distribution to be constructed for different age cohorts and time periods, by gender, ethnicity and so on.

# 4 Simulating Income Profiles

For some purposes – such as the projection of income tax revenues for alternative multi-rate tax functions – it is necessary to model the distribution of income for different ages and cohorts. The previous section models the change in mean log-income with age, $\mu_{t|_c}$, for given cohorts. A convenient extension is to suppose that the income distribution within each age group can be approximated by the lognormal distribution, denoted by $\Lambda\left(\mu_t, \sigma_t^2\right)$, where $\sigma_t^2$ is the variance of logarithms at age $t$. Given, for example, a number of income tax thresholds, it is a simple matter to obtain the proportion of any cohort and age group in the relevant tax brackets. Hence, Subsection 4.1 examines this variance. Subsection 4.2 then illustrates a number of resulting profiles.

## 4.1 Age and the Dispersion of Log-Income

The form of the variation with age in the dispersion of log-income arises from a complex process, involving the pattern of differential income growth among individuals. For example, in the simple unrealistic case where individuals of a given age share equal proportional income changes from one year to the next, there is no change in relative incomes and the variance of log-income remains constant. If there are apparently random variations in growth rates, imposed simultaneously with systematic equalising changes (such that those in lower-income groups experience, on average, relatively higher growth rates), changes in the variance depend on the strength of these component influences. Furthermore, the dynamic process may itself vary over the life cycle of a cohort. Rather than considering these structural factors, the following discussion simply imposes a 'reduced form' specification.

A first indication of the shape of variance profiles with age is provided by the examples shown in Figures 7. This displays, for all males and females, the profiles of the variance of log-income with age, for a range of cohorts. These clearly show more year-to-year variability than the profiles of mean log-income. In addition, the effect of calendar time seems more complex than in the case of mean log-income. Similar graphs are found for the various demographic groups examined above. Unlike the 'time effect' in the context of the mean log-income profile, where 'overtaking' can be related to factors such as overall productivity growth, such obvious interpretations do not arise in the context of the variance of logarithms. Hence, estimates of the broad relationship between the variance and age, reported below, do not contain a term in calendar time.

Not surprisingly, there is a sharp fall at the 'standard retirement age' of 65. Examination of the profiles in the figures suggests that a quadratic is not capable of dealing with the variation. For this reason a cubic was chosen.[10] Considering only ages 20 to 64, Table 3 reports the results of fitting a cubic relationship between the variance of log-income, $\sigma_t^2$, at age, $t$, and age. Not surprisingly, the goodness of fit is generally lower than for estimates of the profiles of mean log-income.

## 4.2 Alternative Age-Income Profiles

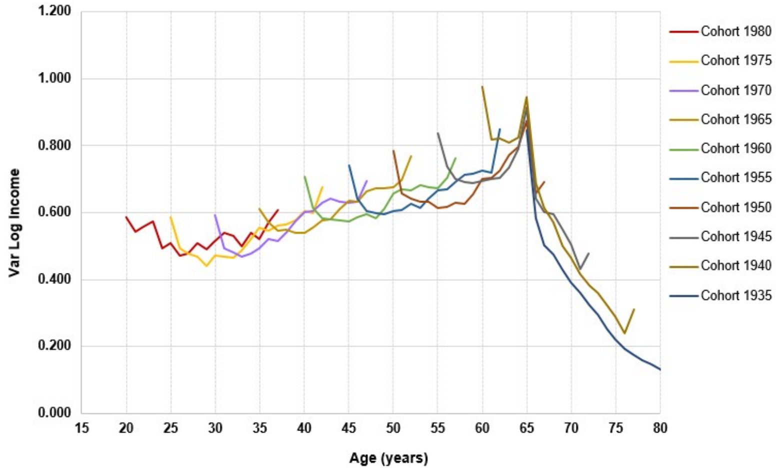
Given the estimates of the profiles of $\mu_t$ (for a given cohort) and $\sigma_t^2$, and the pragmatic assumption that incomes are lognormally distributed as $\Lambda\left(\mu_t, \sigma_t^2\right)$, other profiles can easily be obtained. The median is simply $\exp(\mu_t)$. Furthermore, from the properties of the lognormal, the arithmetic mean income at age $t$, $\bar{x}_t$, is given by:[11]

$$\bar{x}_t = \exp\left(\mu_t + 0.5\sigma_t^2\right) \tag{3}$$

[10] A cubic was earlier found to be appropriate, using more limited NZ data, by Creedy (1997, p. 110), who also found no 'time effects'.

[11] See Aitchison and Brown (1957).

All Males




All Females

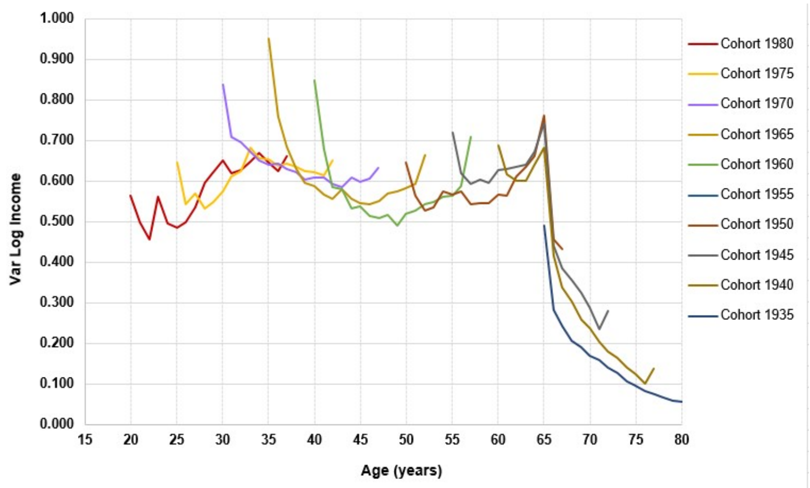



Figure 7: Age and Variance of Log-Income: All Males and All Females

Table 3: Age Profiles of Variance of log-income: Males and Females

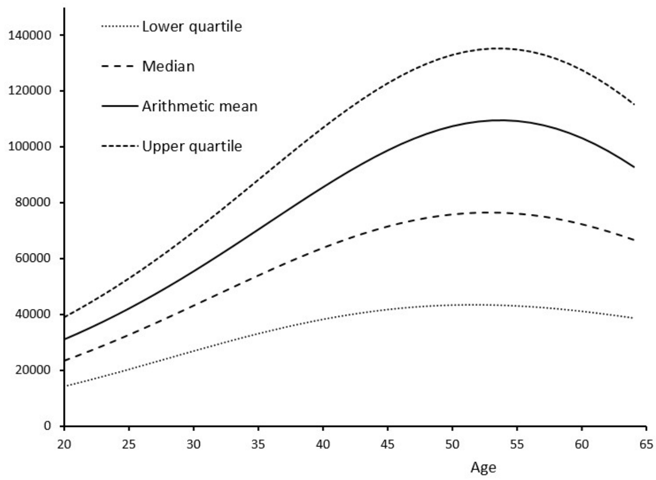
| | Male: All | | Female: All | |
|---|---|---|---|---|
| Coefficient | Estimate | t-value | Estimate | t-value |
| C | 0.5800481 | 23.35 | 0.5800481 | 23.35 |
| Age | -0.0197904 | -6.60 | -0.0197904 | -6.60 |
| $Age^2$ | 0.0013972 | 13.14 | 0.0013972 | 13.14 |
| $Age^3$ | -0.0000205 | -18.18 | -0.0000205 | -18.18 |
| $R^2$ | 0.7397 | | 0.7397 | |
| | Male Maori | | Female Maori | |
| C | 0.4752288 | 22.39 | 0.3823937 | 19.64 |
| Age | 0.0015842 | 0.62 | 0.0079170 | 3.37 |
| $Age^2$ | 0.0003586 | 3.95 | 0.0000070 | 0.08 |
| $Age^3$ | -0.00000836 | -8.68 | -0.0000042 | -4.76 |
| $R^2$ | 0.7329 | | 0.7872 | |
| | Male Non-M + Non-P | | Female Non-M + Non-P | |
| C | 0.7884353 | 30.33 | 0.766025 | 36.45 |
| Age | -0.0363858 | -11.60 | -0.015658 | -6.18 |
| $Age^2$ | 0.0020925 | 18.81 | 0.000769 | 8.55 |
| $Age^3$ | -0.0000286 | -24.21 | -0.000013 | -13.15 |
| $R^2$ | 0.4243 | | 0.5647 | |
| | Male School | | Female School | |
| C | 0.5551574 | 19.62 | 0.626723 | 19.16 |
| Age | -0.0266551 | -7.81 | -0.012474 | -3.16 |
| $Age^2$ | 0.0015873 | 13.11 | 0.000634 | 4.53 |
| $Age^3$ | -0.0000214 | -16.69 | -0.000010 | -6.86 |
| $R^2$ | 0.6282 | | 0.6004 | |
| | Male Post School | | Female Post School | |
| C | 0.578961 | 21.81 | 0.429078 | 17.65 |
| Age | -0.0344597 | -10.76 | -0.015780 | -3.59 |
| $Age^2$ | 0.0017829 | 15.69 | 0.000845 | 5.42 |
| $Age^3$ | -0.0000228 | -18.94 | -0.000013 | -7.73 |
| $R^2$ | 0.6251 | | 0.5235 | |
| | Male University Degree | | Female University Degree | |
| C | 1.019419 | 26.60 | 0.925605 | 19.51 |
| Age | -0.067589 | -14.62 | -0.035053 | -6.13 |
| $Age^2$ | 0.0029378 | 17.91 | 0.001355 | 6.67 |
| $Age^3$ | -0.0000335 | -19.28 | -0.000017 | -7.75 |
| $R^2$ | 0.5511 | | 0.3767 | |

Using the fact that the quartiles of the standard Normal distribution are $\pm 0.67448$, quartiles of the income distribution at age $t$ are $\exp(\mu_t \pm 0.67448\sigma_t)$.

Figure 8 shows projected income profiles for cohorts of all males and all females. In each case the cohort is of those aged 20 in the year 2000. In making such projections it is also a simple matter to change assumptions about future productivity growth, by varying the parameter, $\beta$, in the profile for $\mu_{t|c}$. Projections for other cohorts can again easily be made by suitable variation of the value of $c$. For example, the cohort aged 20 in 1980 would have a value of $c$ of -19.5. Compared with projections in Figure 8, the profiles would be both lower and flatter, since the constant term and the coefficient on $t$ in $\mu_{t|c}$ would be lower.

Examples of projections of different profiles for different qualification levels are shown in Figure 9, for those with a school qualification only, and Figure 10, for those with a university degree. In the latter case the vertical axis has been extended to allow for higher incomes. These diagrams clearly illustrate the fact that higher education levels produce more 'peaked' age-income profiles. It may initially be tempting to compute the present value of projected incomes along the 'arithmetic mean' profile, and refer to this as a measure of arithmetic mean 'career income'. However, the latter cannot be constructed independently of the way in which individuals move within the distribution of their cohort from year to year.

All Males




All Females

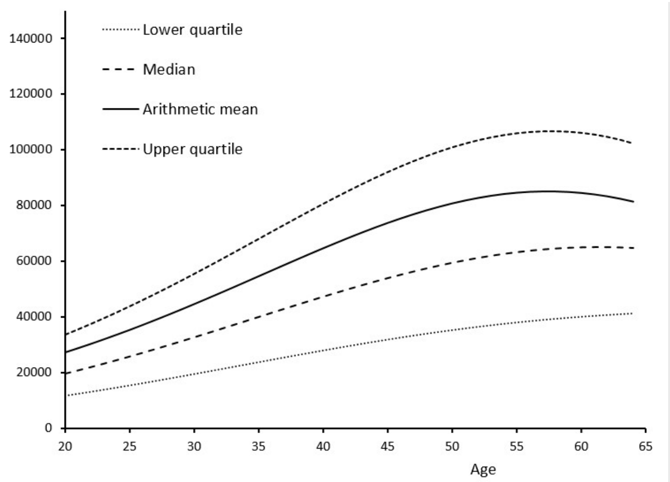
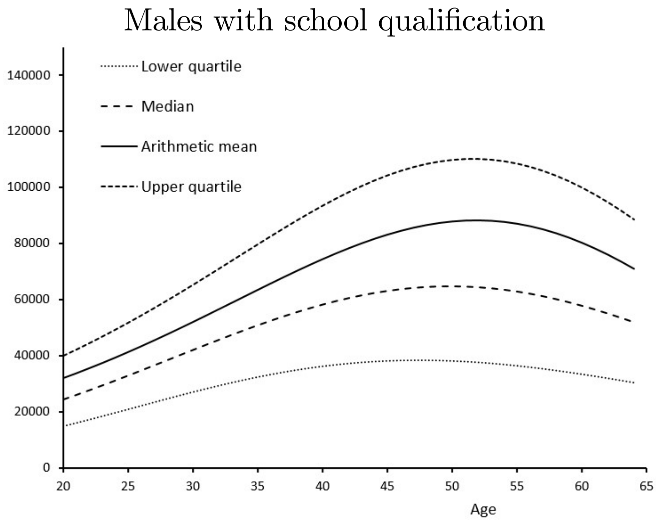



Figure 8: Alternative Projected Profiles for Cohorts Aged 20 in 2000: All Males and All Females



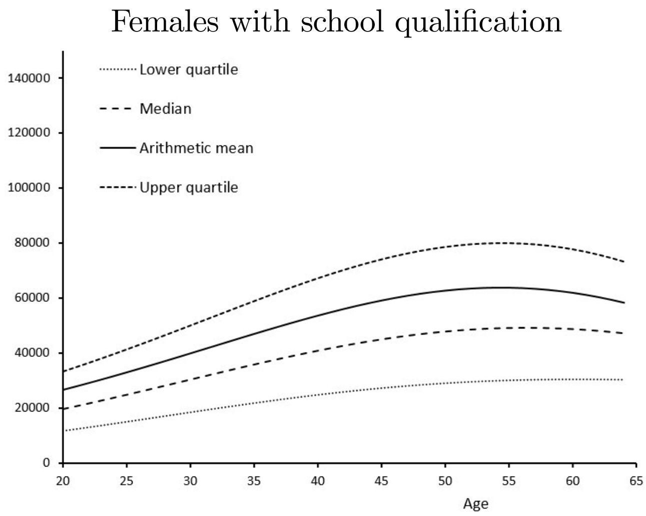




Figure 9: Alternative Projected Profiles for Cohorts Aged 20 in 2000: School Qualification Only

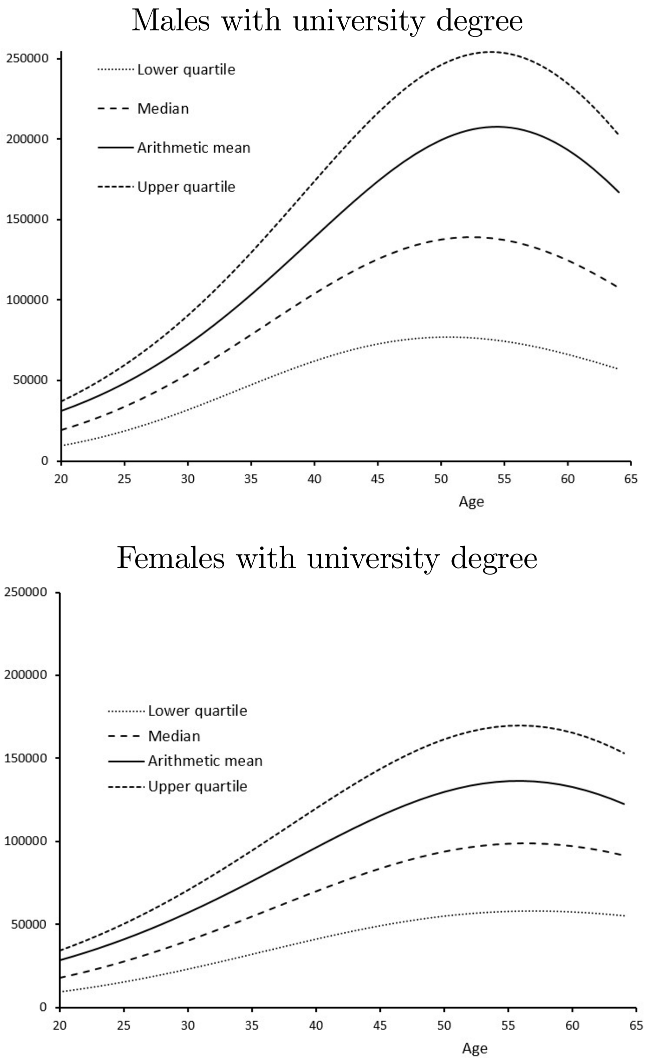



Figure 10: Alternative Projected Profiles for Cohorts Aged 20 in 2000: University Degree

# 5 Conclusions

This paper has examined age-income profiles of individual taxpayers in various demographic groups in New Zealand, using information from a special longitudinal dataset on annual taxable incomes over an 18-year period from 2000 to 2017, compiled using the Statistics New Zealand Integrated Data Infrastructure. The experience of separate cohorts, over different parts of the life cycle, were compared, while linking Inland Revenue data with other administrative data sources, enabled separate profiles for various demographic groups to be constructed.

Summary information, in the form of diagrammatic illustrations of age-income profiles, confirmed a number of gender, ethnicity and educational differences in the age- and calendar time-related properties of taxable incomes in New Zealand. Namely, that while male and female incomes both tend to rise with age, male incomes are higher on average than female incomes at each age, reach a maximum level at a younger age, and both genders experience 'overtaking' by later cohorts of their earlier equivalents.

The form of the profiles, for a variety of demographic groups, was estimated using a specification involving quadratic age effects and a linear time effect. The time effect is associated with factors such as productivity change, and generates the 'overtaking' effect, despite the absence of explicit cohort effects on the profiles. The variation in the variance of log-income with age was found to be cubic for all groups examined. It was shown how the combination of the age-profiles of the mean and variance of log-income, with an assumption that income in each age group can be approximated by the log-normal distribution, produces a model of the changing distribution of income with age. This permits the projection of alternative quantiles of the distribution for different cohorts and, for example, enables projections of income tax revenue to be made for assumptions about productivity growth and changes in the demographic structure of the population, along with the income tax function.

# References

[1] Aitchison, J. and Brown, J. A. C. (1957) *The Lognormal Distribution.* Cambridge: Cambride University Press.

[2] 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, New Zealand.

[3] Berger, M. (1985) The effect of cohort size on earnings growth: a reconsideration of the evidence. *Journal of Political Economy*, 95, 561-573.

[4] Connolly, R. (1986) A framework for analysing the impact of cohort size on education and labour incomes. *Journal of Human Resources,* 21, 543-562.

[5] Creedy, J. (1985) *Dynamics of Income Distribution.* Oxford: Basil Blackwell.

[6] Creedy, J. (1992) *Income, Inequality and the Life Cycle*. Aldershot: Edward Elgar Publishing.

[7] Creedy, J. (1997) *The Statics and Dynamics of Income Distribution in New Zealand*. Wellington: Institute of Policy Studies.

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

[9] Weiss, Y. and Lillard, L. A. (1978) Experience, vintage and time effects in the growth of earnings, *Journal of Political Economy*, 87, 565-97.

[10] Welch, F. (1979) Effects of cohort size on earnings: the baby boom babies' financial bust, *Journal of Political Economy*, 87, 565-97.

# Appendix: Age-Income Profiles for Taxpayer Groups

This Appendix reports age-income profiles for various taxpayer subgroups based on gender, ethnicity and educational differences. Notice the difference in scale for the vertical axes in some cases.

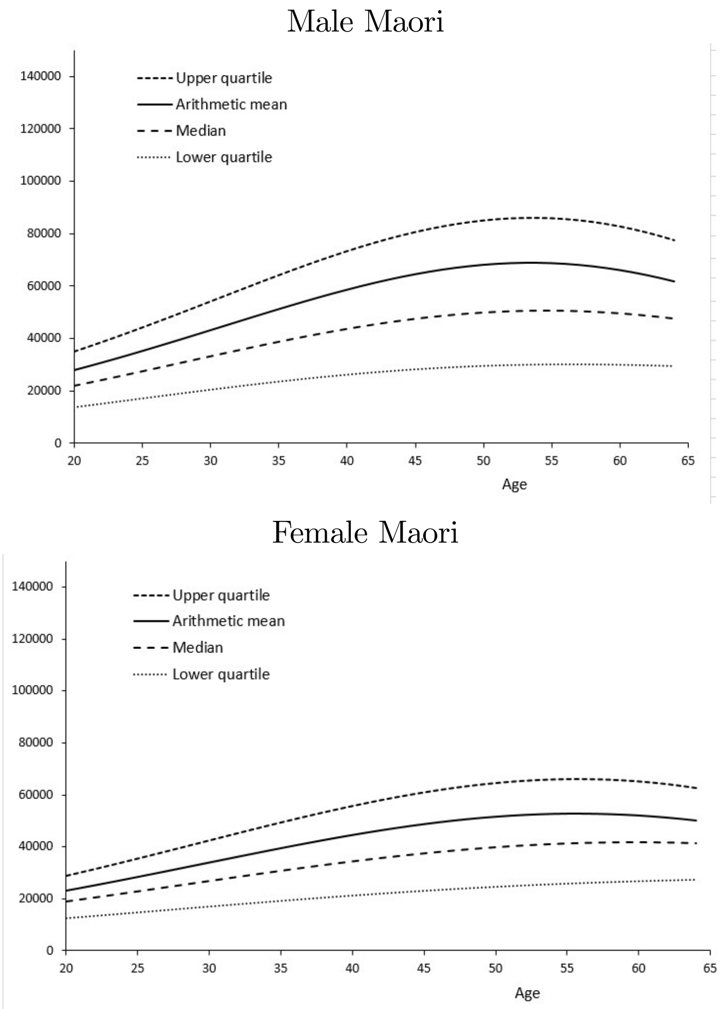



Figure 11: Age-Income Profiles: Maori Subgoup

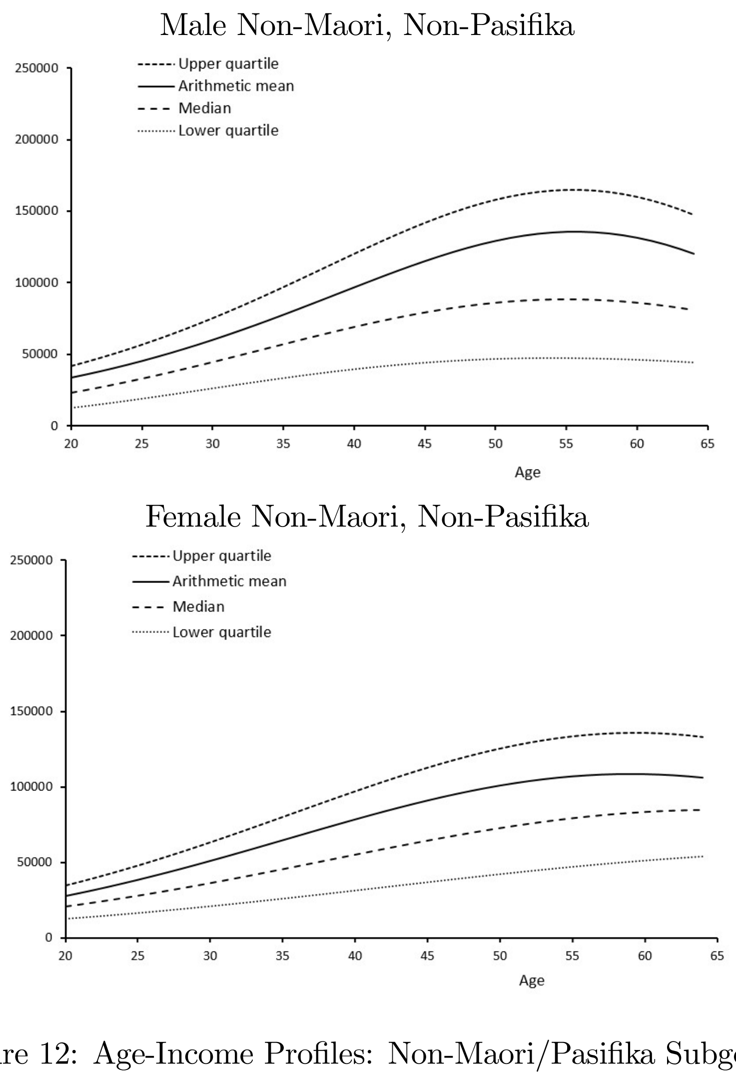
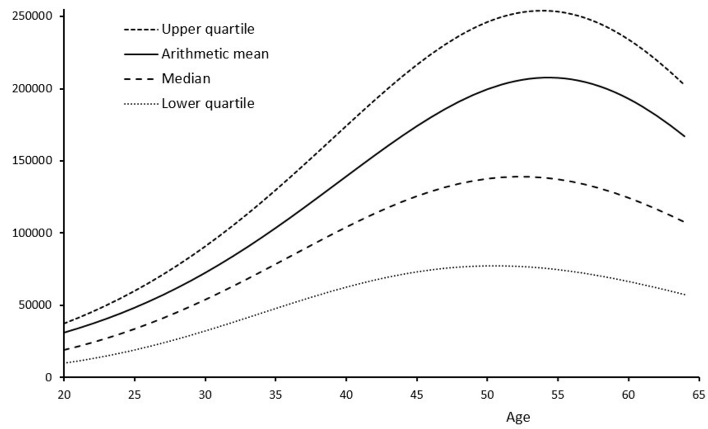
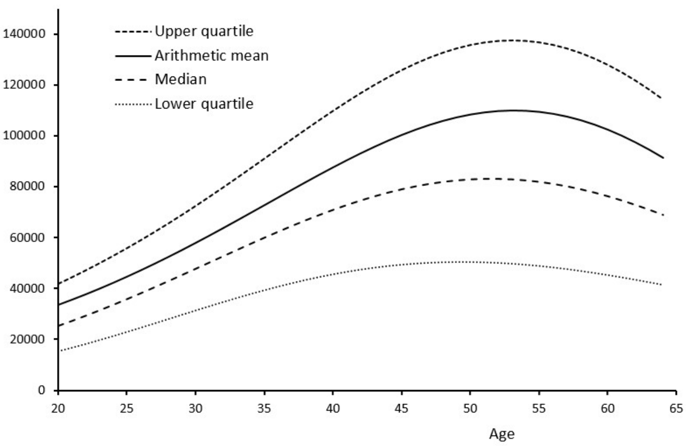



Figure 12: Age-Income Profiles: Non-Maori/Pasifika Subgoup

Male University Qualification




Male Post-School Qualification




Male School Qualification

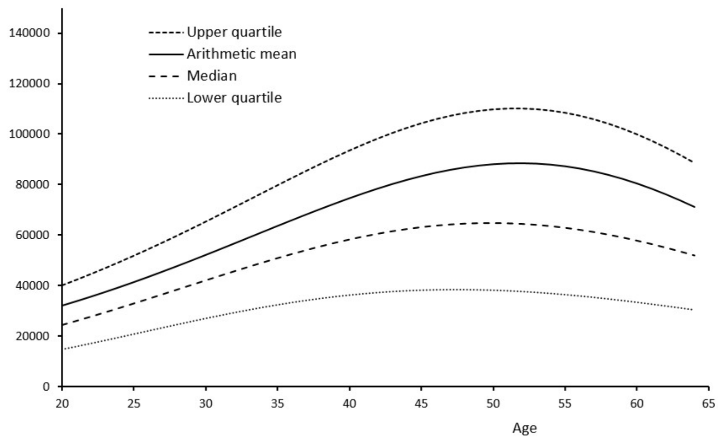



Figure 13: Age-Income Profiles: Male Educational Subgoups

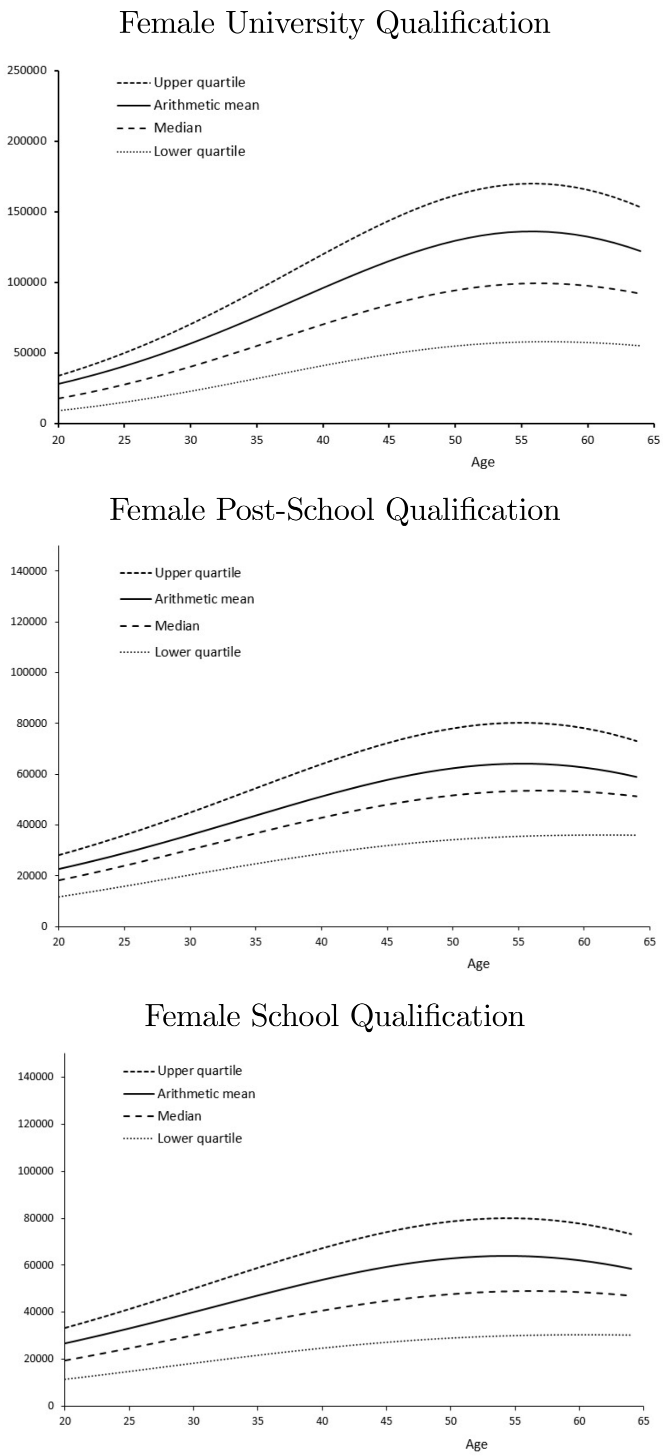



Figure 14: Age-Income Profiles: Female Educational Subgoups

## About the Authors

Nazila Alinaghi is a Research Fellow in Public Finance at Wellington School of Business and Government, Victoria University of Wellington, New Zealand.
Email: nazila.alinaghi@vuw.ac.nz

John Creedy is Professor in Public Economics and Taxation at Wellington School of Business and Government, Victoria University of Wellington, New Zealand.
Email: john.creedy@vuw.ac.nz

Norman Gemmell is the Chair in Public Finance at Wellington School of Business and Government, Victoria University of Wellington, New Zealand.
Email: norman.gemmell@vuw.ac.nz