ISER Working Paper Series

Variations in Earnings Growth: Evidence from Earnings Transitions in the NZ Linked Income Survey

Ron Crawford
Economic Consultant

No. 2009-18
June 2009

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

www.iser.essex.ac.uk

Variations in earnings growth:
Evidence from earnings transitions in the NZ Linked Income Survey

Non-technical summary

This paper uses the Linked Income Supplement (LIS) of the New Zealand Household Labour Force Survey (HLFS) to investigate the annual transitions in hourly earnings of working age individuals over the years 1997 to 2004. The primary aim is to estimate the determinants of annual changes in hourly earnings, for those who have positive hourly earnings in each year.

A first issue that needs to be addressed is the significant attrition in the LIS mainly due to the fact that the HLFS only follows individuals who remain at the same address between years. A second more important issue is the possibility of biased estimates due to focusing only on people who have positive hourly earnings in each year. Because the LIS provides only two observations at a one-year interval on the earnings of individuals, it is not possible to use sophisticated statistical methods to account for these possible sources of bias. Instead, the paper shows that, in practice there are only small differences in the observed characteristics of people in this study, compared with people of the same age represented in the HLFS who have positive earnings in any year. The shape of the earnings distribution is also quite similar for both groups.

The paper shows that the probabilities of moving across earnings quintiles between years in the LIS data are quite similar to those found in earlier analyses using tax data and show a moderate degree of earnings mobility in New Zealand. Differences can potentially be accounted for by known differences in the earnings measure used in the two sets of analyses.

The paper finds that differences in educational qualifications are associated with differences in earnings growth between years. Better educated people are likely to have better learning ability and better opportunities to learn while in employment. However, because of data limitations, the results are best interpreted as associations rather than as necessarily showing cause and effect.

Variations in earnings growth:
Evidence from earnings transitions in the NZ Linked Income Survey

Ron Crawford

Economic consultant

June 2009

Abstract

This paper uses the New Zealand Linked Income Supplement (LIS) to investigate the annual transitions in hourly earnings of working age individuals over the years 1997 to 2004. I first construct transition matrices for annual changes in weekly and hourly earnings, to enable comparison with previous analyses using New Zealand tax data. I then estimate the determinants of annual changes in hourly earnings using OLS and quantile regressions. Differences in human capital are associated with differences in the rate of earnings growth. The results were broadly similar across the sub-periods 1997-2001 and 2001-2004.

Keywords: Earnings dynamics; earnings growth; human capital

Correspondence: rjcrawf@xtra.co.nz

Acknowledgments

I thank Dave Maré and Stephen Jenkins for ongoing advice and support during the progress of this study, and Dean Hyslop and Steve Stillman for helpful comments on later drafts. I thank my previous employer the New Zealand Treasury for leave to begin this work at the University of Essex, the Institute for Social and Economic Research for hosting me, and Motu Economic and Public Policy Research for financial support to complete this study. Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the author, not Statistics New Zealand.

I Introduction

This paper uses the Linked Income Supplement (LIS) of the New Zealand Household Labour Force Survey (HLFS) to investigate the annual transitions in hourly earnings of working age individuals over the years 1997 to 2004. The paper looks at movements up and down the earnings distribution and the demographic and human capital characteristics associated with those movements. The work complements the household income based analysis of child poverty dynamics reported in Ballantyne *et al*., (2003) by looking at an aspect of the evolution of individual earnings.

There are many possible motivations for investigating earnings dynamics, and thus many different possible research strategies. For instance, household members entering or leaving the labour market, or experiencing periods of unemployment can have dramatic effects on household incomes. The current paper does not investigate these events because they are best placed in a household or economic family context as in Ballantyne *et al*., (2003). Similarly, the earnings outcomes of the interaction between personal characteristics and characteristics of jobs is best studied using specially constructed longitudinal datasets that link data on individuals with data on successive jobs.[1]

This paper instead focuses on differences in earnings growth associated with demographic and educational differences. Studies based on both cross-sectional and longitudinal data show that the age-earnings profiles of people with given demographic characteristics vary in their shape. In particular, the profile for those with higher levels of education rises more steeply, attains higher levels and declines later than for others.[2]

While these sources of growth in earnings generally have a small annual effect, compared to other labour market events, the effects may persist and accumulate over

[1]Statistics New Zealand has constructed such a database from longitudinal employer tax data sourced from the Department of Inland Revenue and business data from the Statistics New Zealand Business Frame. The data is available for analysis subject to strict privacy and confidentiality protocols http://www.stats.govt.nz/products-and-services/table-builder/leed-table-builder.htm .

[2] See, for instance, Huggett *et al*. (2007) who used longitudinal data from the Michigan Panel Study of Income Dynamics (PSID). Maani (2000) presents cross-sectional age income profiles by educational level using data from New Zealand censuses.

time, and thus mark out broad population differences in hourly earnings trajectories. Small persistent differences in the evolution of earnings year by year can lead to large differences in life-time income and wealth.[3] Understanding the enduring characteristics associated with differences in the evolution of earnings, and whether their effect changes over different time periods, may thus be important for understanding life-time differences in income and wealth.

The HLFS LIS provides weekly and hourly earnings measures, as well as a banded annual earnings measure. Because it is banded, the annual measure is not suitable for the analyses presented in this paper. An argument could be made for either weekly or hourly earnings to better reflect the effect of stable personal characteristics on earnings transitions. The hourly earnings measure comprises a combination of directly reported hourly wages by those paid an hourly wage, and a derived hourly wage by those receiving a salary and reporting their current or usual hours of work. In practice, regression results using weekly and hourly earnings were generally qualitatively similar, and the hourly earnings measure was chosen for the main analyses.[4]

The paper focuses in particular on individuals with positive hourly earnings. Most of the findings reported are conditional on individuals having positive hourly earnings at each of the two points of time at which their earnings are observed.

The rest of the paper is structured as follows. Section 2 briefly reviews the relevant literature, while section 3 focuses on New Zealand research on earnings dynamics, and the data on which it is based. Section 4 briefly describes the HLFS LIS data. Section 5 discusses the methodology adopted. It then describes the sample

[3] Consistent with this, Baker (1997) finds evidence for heterogeneous growth rates in annual earnings, with individuals one standard deviation above the mean enjoying a 20-30 percent earnings advantage in ten years.

[4] "Usual total hourly earnings" in the first wage and salary job is used for the main analysis. This is a composite of self-reported hourly earnings for those paid by the hour, and usual total weekly earnings divided by usual total hours of work for those receiving salaries. Unpublished Department of Labour analyses show that just over a third of individuals in the 2005 HLFS-Income Supplement (from which the LIS is derived) reported an hourly wage rate. The incidence of such individuals falls to about a quarter at the bottom of the distribution, below the minimum wage (personal communication from Jason Timmins, Department of Labour). This is, apparently because many salary earners report considerably higher than the average usual hours of work, leading to their calculated hourly earnings rate being low. Atkinson *et al* (1992) discuss systematic differences in hourly earnings measures based on directly stated hourly rates, and on an hourly rate calculated from salaries and hours worked.

characteristics of the longitudinally-weighted sample and compares these to HLFS cross-sectional data for the same period. Section 6 reports the findings. It first reports estimates of one-year transition probabilities for weekly earnings, and compares these with individual annual market earnings transitions estimated from tax data in Hyslop (2002). It then reports one-year hourly earnings transitions and compares these with the results for weekly earnings. Finally, it presents the results of regressions with the change in log hourly earnings as the dependent variable. The main analyses are repeated for sub-periods. Section 7 discusses the findings and concludes.

2 Research on earnings dynamics

Empirical studies of earnings dynamics have a long history, and have been motivated by a range of economic issues. Initially these included the design of pension schemes and income tax regimes where earnings mobility could have a significant influence on the viability and equity of arrangements. This led naturally to attention to the distribution of life-time earnings and an interest in the relationship between earnings mobility and labour market institutions (Atkinson *et al.*, 1992). More recently there has been a focus on the relationship between cross-sectional earnings inequality and life-time earnings inequality, trends in cross-sectional earnings inequality, explaining life-time patterns of consumption inequality in terms of differences in earnings risk[5] and identifying differences in intergenerational correlations in earnings inequality associated with using cross-sectional and life-time measures.[6]

While there are many strands to this literature, one recent focus has been whether increasing cross-sectional earnings inequality can best be explained by increasing transitory earnings mobility, or an increase in permanent inequality (or a combination of both). Even allowing for differences in definitions, types of data, and methodologies, studies appears to have found different answers in different time periods and in different countries.[7] Over the last decade models of the evolution of earnings (and consumption) have become increasingly complex, and have utilized

[5] See, for instance, Guvenen (2007).
[6] Solon (1999).
[7] See, for instance, Haider (2001) for the United States; Baker and Solon (2003) for Canada, Burkhauser *et al.* (1997) for the United States and Germany.

simulation methods to estimate model parameters. The more complex models have utilised very long longitudinal datasets,[8] and sometimes very large datasets[9].

A range of evidence shows that individuals face heterogeneous earnings growth profiles over their life-time,[10] and that this is partly associated with differences in education levels and other observable differences,[11] and partly associated with unobserved differences.[12] Altonji *et al.* (2009) develop a complex model that allows for heterogeneity in growth rates of earnings associated with initial education and unobserved factors, and incorporates changes in employment. They calibrate model parameters through simulation techniques using data on the earnings of male heads of households in the Michigan Panel Study of Income Dynamics (PSID) over the years 1975 to 1996. They find that education accounts for about one third of the variance in lifetime earnings, while permanent unobserved heterogeneity accounts for a further 11 percent.[13]

Huggett *et al.* (2002, 2007) also investigate the role of human capital and unobserved differences in learning ability in explaining the variation in lifetime earnings. In the first paper they show that a human capital model can replicate salient properties of the distribution of earnings over the working-life cycle for a typical cohort as the cohort ages, and that differences in learning ability are essential to produce an increase in earnings dispersion over the life-cycle and account for most of the variation in life-time earnings. In the second paper they use a simulation approach to show, using PSID earnings data on male heads of households over the years 1969-2004, that initial conditions at age 20, including human capital and learning ability, account for more of the variation in life-time earnings than do differences in shocks received over the

[8] Many studies have used the PSID, usually restricting analysis to working age men.
[9] Baker and Solon (2003) use a large Canadian tax dataset that allows a particularly complex model to be studied.
[10] Baker (1997), for instance, finds evidence for heterogeneous growth rates in annual earnings, with individuals one standard deviation above the mean enjoying a 20-30 percent earnings advantage in ten years. For other references see Baker (1997) and Baker and Solon (1993).
[11] Haider (2001), using the PSID for the years 1967-1991, finds that one third of the persistent component of increased inequality in annual earnings is attributable to changing returns to education.
[12] Guvenen (2007) finds that a model of the evolution of earnings and consumption that relies on heterogeneous growth profiles rather than one that relies on large and persistent heterogeneous shocks can better explain the non-concave shape of the age-earnings inequality profile and the fact that consumption profiles are steeper for higher educated individuals.
[13] Geweke and Keane (2000), using the PSID for the years 1968-1989, also find support for unobserved individual heterogeneity accounting for the majority of life time variance in earnings.

lifetime. Moreover they find that human capital and learning ability at age 20 are correlated (individuals with greater learning ability devote more time to acquiring human capital). In their model, human capital raises an agent's earnings profile, while learning ability produces a steeper profile. If data on learning ability is unavailable in simple models of earnings growth, then human capital variables may, instead, capture the earnings growth effect.

3 Research on earnings dynamics in New Zealand[14]

To date, three data sources have been used for the analysis of income dynamics in New Zealand – tax data held by the Inland Revenue Department (IRD), data on family income over a 14-year period for children in the Christchurch Health and Development Study (CHDS)[15] and, recently, the LIS. The Survey of Family, Income and Employment (SoFIE), now (in 2009) in its sixth year in the field, is beginning to provide a purpose built resource that could be used for analyses of income dynamics as data is made available by Statistics New Zealand.[16] Of these data sources only the tax data have been used for the analysis of the earnings dynamics of individuals.

Data from individual income tax records

The New Zealand Inland Revenue Department (IRD) has constructed a longitudinal database based on tax returns for a representative sample of individual tax payers for various periods over the last thirty years. Hyslop (2000a) provides a careful description and analysis of the quality of the current longitudinal data set. Three substantial sets of analyses have been carried out using this data.

Smith and Templeton (1990) used IRD data for the years 1979 through 1987 to construct multi-year quintile group transition matrices, and to study transitions over time in the principal source of earnings. They also estimate parameters for a simple

[14] O'Dea (2000) provides an extensive review of research in New Zealand to that point on the distribution of income and income dynamics.
[15] See Maloney and Barker (1999), Maloney (2000) for studies of the family income dynamics of children in the CHDS. Ballantyne *et al.* (2003) provide a brief description of this research.
[16] Currently data from the first four waves are available for analysis.

model of annual earnings change (dependent on the level of earnings in the previous year). They find a familiar pattern of decreasing probability over successive years of remaining in the original earnings quintile group, but, at the same time, a much higher probability of remaining in the group with the highest earnings. A proportionately larger change in probabilities occurs between the first and second year, compared with subsequent years.

Creedy (1997) uses IRD data for 1991-1993 to estimate a model of relative taxable income[17] mobility that depends on three factors: regression towards the mean, the degree of serial correlation in successive relative changes in income, and ‘chance’ variation. He uses the estimated parameters to simulate income dynamics over the life cycle. He finds that there is some regression towards the mean, for both men and women, but this varies with age. For men (and for younger women) there is some negative serial correlation in relative proportional income changes from year to year. There is also a substantial amount of apparently ‘random’ relative movement from year to year. These factors generate a changing dispersion of annual income over the life cycle – for men, rising sharply in the first few years, then falling and again gradually rising until reaching a maximum around 60 years of age.[18] Creedy goes on to use this model to analyse the effects of various New Zealand income tax regimes on inequality measured over different periods of time.

Hyslop (2000a & b) uses IRD data for the four-year period 1994-97.[19] In the first paper, he constructs multi-year transition probability matrices, separately for men and women, and adjusting for age (and year effects), using a measure of annual market income. These matrices, even with adjustment for age and other methodological differences, show a similar pattern to those of Smith and Templeton (1990), though for men the probability of remaining in the original quintile group was substantially higher in the lowest group, and somewhat lower, in the highest group.[20]

[17] He focuses on individuals who derive their main source of income from wages and salaries, but also includes some income from welfare transfer payments.
[18] Creedy separately shows that there is a complex relationship between income mobility, the age structure of a population, and cross-sectional and life cycle patterns of inequality
[19] Later extended to 1998 in an unpublished analysis (Hyslop, 2002). This analysis, kindly provided by the author, is used in the current paper to compare results using LIS and tax data.
[20] For women the probability of remaining was substantially higher in the lowest group, lower in the next and much the same in the remaining groups. But there are significant differences between the studies in construction of quintile groups and also a higher probability of a transition to missing data in

Hyslop (2000a) also analyses the covariance structure of logarithms of annual income. He finds that a large fraction of the observed differences in income are transitory – less than 50% (but about 60% when outliers are excluded)[21] of the differences in income persist (as measured by the correlations) after three years.

Hyslop (2000b) uses the same four years of data to investigate the effect of welfare benefit receipt on subsequent market income and gross (market plus benefit) income levels. Using lagged income values, and indicators for contemporaneous and lagged benefit receipt he controls for heterogeneous differences in individual incomes, spurious effects of contemporaneous benefit receipt and possible endogeneity with incomes. He found no systematic evidence of a positive or negative effect of benefit receipt on incomes – though the results are generally imprecisely measured, and sensitive to choice of specification.

These analyses illustrate both the strengths and weaknesses of the IRD data set for the analysis of income dynamics. First, the sample is large – for instance the transition probabilities in Hyslop (2000a) were estimated for a sample size of 23,145 individuals. Second, it is possible to construct relatively long panels – in the case of Smith and Templeton (1990) eight years. This allows depiction of patterns of income receipt over time – with obvious improvements in relevance for understanding effects on well-being. In addition, observations of the same individuals over a number of years allow econometric issues to be addressed with greater sophistication. Creedy (for instance) required three periods of data to estimate the degree of serial correlation in successive relative changes in income. The specifications estimated in Hyslop (2000b) need four periods of data to address issues of unobserved heterogeneity, and endogeneity in changes in benefit receipt and income levels, as well as to test the model specification.

On the other hand, the IRD data contains very little information on individuals - essentially it is restricted to age, gender, and annual taxable income, to some extent

Smith and Templeton (1990) possibly because of the different ways in which the data sets were constructed.

[21] Hyslop notes that the figure of 60% is close to estimates using panel data in the United States.

disaggregated by source. Thus it is not possible to control explicitly for a range of factors that may influence income dynamics, nor, except to a very limited extent, investigate the determinants of transitions. In addition, the sample is subject to selection bias and non-random attrition[22] (low income earners are not required to file tax returns) and to error in the measurement of income.[23]

The Linked Income Supplement of the Household Labour Force Survey

The New Zealand HLFS has an eight-quarter rotating panel design. Since 1997, an income questionnaire has been administered in the June quarter. For each panel there are, therefore, two successive annual observations on income, allowing the study of one-year transitions. (A fuller description of this data set, and its limitations, is provided in Section 3).

Ballantyne *et al.* (2003) are the first to use this data for the study of income dynamics – in this case child poverty transitions. They used data for 1997-2000. They compare New Zealand with five other countries covered by Bradbury *et al* (2001). Overall, New Zealand appears to have comparatively high rates of relative child poverty with 13 percent of children living in poverty with the poverty threshold set at 50 per cent of median equivalised household income. But New Zealand also has high rates of transition into and out of poverty – with 8.2 percent of those not in poverty entering poverty, and 59 percent of those in poverty exiting each year.

They also replicate Jenkins and Schluter's (2003) analysis for Britain and West Germany of the probability of trigger events that are associated with transitions, and, conditional on these events, the probability that a transition is made. They find that children in lone parent families are much less likely to exit poverty in New Zealand compared with Britain and West Germany. They are also much less likely to exit poverty, given that a trigger event (such as parental re-partnering or the household gaining a full-time worker) has occurred. On the other hand, conditional on these

[22] Smith and Templeton (1990), Hyslop (2000a).
[23] Hyslop (2000a) investigates this issue carefully, by comparing information from different sources within the IRD data. He concludes that there is a high degree of consistency in the data, although inconsistencies point to the need for care in the handling of outliers.

events, New Zealand children in families with two parents present are more likely to exit poverty than in the other two countries. Given an economically significant rise in labour earnings, children in New Zealand are more likely to exit poverty, than in Britain and West Germany.

An important limitation of the LIS data is that the survey is based on geographic addresses, and does not follow individuals (or groups) who leave the address after the first wave. In addition Ballantyne *et al.* (2003) restrict their estimates to children in households in which all eligible adults provide full information on income. These factors lead to high rates of non-random attrition from the sample (with 38.5 percent of children in poverty, and 33.1 percent of those not in poverty being dropped from the sample between waves). The authors model attrition probabilities using observed characteristics, and then predict transition probabilities for those subject to attrition. They conclude that attrition has only modest effects on estimated transition probabilities.

Survey of Family, Income and Employment (SoFIE)

In October 2002, Statistics New Zealand began the first wave of data collection for SoFIE. This is an eight-year longitudinal survey with families living in 10,000 randomly chosen households in the first wave as the focus of analysis. The design follows adults (those over 15 years) over the eight annual waves, collecting a range of core data on family structure, labour market status and income. Other data are being collected less frequently. The survey is designed to facilitate the study of family and individual income dynamics. Currently (2009) data from the first four waves is available for analysis.

Summary

Research on earnings dynamics in New Zealand has been severely constrained. Tax data provides a limited range of covariates; the CHDS covers a specific cohort of children born in a particular location and uses only a banded measure of family annual income; and the LIS provides data for one-year transitions only. All three sets of data are subject to significant attrition, and while estimates can be adjusted on the basis of observed characteristics, it is less easy to take into account biases due to unobserved factors correlated both with the probability of attrition, and the outcome being studied. Nevertheless, research to date suggest that there is a moderate to high rate of individual earnings mobility in New Zealand.

It is expected that Statistics New Zealand will make longer periods of data from SoFIE available over successive years. In the meantime, analysis of individual earnings dynamics using the LIS can provide some information on the nature of annual transitions, and correlates of changes in earnings. In addition, the HLFS-IS on which the LIS is based is scheduled to continue year-by-year, whereas the last wave of SoFIE will conclude in 2010.

The next section looks more closely at the characteristics of the LIS and its strengths and weaknesses for the analysis of individual income dynamics.

4 The HLFS LIS data

The HLFS is a quarterly clustered random survey currently covering about 15,000 dwellings and 30,000 people.[24] The survey oversamples clusters with higher proportions of ethnic minority populations. The design comprises a rotating panel, with each panel lasting eight quarters.[25] The Income Supplement (IS) occurs annually every June quarter, providing the opportunity to link income data for the same individuals across a one-year span, to form the LIS. The IS collects information on current and usual hourly and weekly earnings and hours of work, and income from self-employment and other private sources, and from government benefits over the reference week of the HLFS. It also collects retrospective information on annual personal income from all sources by 13 ranges, over the previous year.

[24] Ballantyne *et al.* (2003) provide a detailed and extended description of the LIS and its strengths and weaknesses for the analysis of income dynamics.
[25] For a period in 1998/99 and 1999/2000, one quarter rather than one eighth of the sample was rotated out each quarter.

The sampling frame selects a representative address, and at each quarter collects information focusing on people usually resident there at the time of the interview. A household questionnaire collects demographic data and information on relationships among household members. Each quarter a personal questionnaire on labour market issues is administered in relation to everyone of working age (15 years and above) who is in the scope of the survey. Proxy responses on behalf of relatives are permitted. A questionnaire is not completed in relation to anyone who, though usually resident, is away from the address for more than six weeks. Anyone who has permanently left the address after the first interview is not followed; information is, on the other hand, collected for people who have become usually resident at the address since the first interview.

In the June quarter a questionnaire on income is administered in addition to the personal questionnaire. Proxy responses are not permitted. Response rates to the IS are lower than for the HLFS, and contemporaneous information from the latter may be used by Statistics New Zealand to impute income data. The method uses demographic characteristics to match records and assigns income data from the first matching record. *A priori* it is likely that imputation will add a spurious source of volatility in incomes from one year to the next.[26] For this reason income dynamics analyses that exclude records with imputed incomes are preferred.[27]

A combination of geographic mobility and incomplete or no response to the IS leads to high rates of attrition from the sample that may produce significant biases in estimates. This study uses data from seven successive one-year panels collected between June 1997 and June 2004. The LIS links this data for individuals for whom complete information is available in each of the two waves. There are two versions – one of which includes and the other excludes records with imputed incomes. Statistics New Zealand provides alternative weights for the resulting samples that reflect the probability of a household being selected, and that make adjustments for non response to the HLFS, IS attrition, and to reflect the gender composition of the

[26] Consistent with this, a preliminary analysis showed that the probability of movement across earnings quintiles is higher if imputed incomes are included.
[27] About 20 percent of individuals in the LIS have incomes imputed in one or both years.

population from other data. The current research uses the sample without imputed incomes.

5 Research strategy, estimation issues and methodology

Attrition and sample selection

A first obvious issue arises from the nature of the HLFS sampling frame. This follows household members over time, only while they continue to live at a particular address. Significant attrition occurs not only because of non-response, but also simply because individuals leave addresses between waves. Preliminary analyses showed that this attrition is correlated with factors that are also associated with differences in earnings changes between years. Nevertheless the income dynamics literature sometimes finds that adjusting transition estimates for attrition, based on observable characteristics, makes a negligible difference to estimates.[28] This is likely in part because observable characteristics explain very little of the individual variation in income changes.

A second selection issue arises from limiting the main analyses to those with positive hourly earnings in each wave of the linked data. Again those who are dropped from the analysis are likely to differ in some important respects from those who remain, and on dimensions that are associated with earnings changes between years.

As with attrition, adjustment to estimates for differences in observable characteristics could be modelled, but a more important issue is likely to be differences due to unobservable characteristics. Because there are only two periods of data, it is not

[28] Ballantyne *et al.* (2003) follow such modelling approach to account for the effect of attrition on their for estimates of child poverty entry and exit using the HLFS LIS. They conclude that “attrition bias is not likely to significantly impact on our results.” Fitzgerald *et al.* (1998) find that the large amount of attrition in the Michigan Panel Study of Income Dynamics does not appear to have led to serious distortions in the representativeness of the survey. In particular they find that the selection that occurs is based on transitory components that fade over time and are moderated by regression-to-the-mean effects. Cappellari and Jenkins (2008) show that sample attrition is relatively unimportant compared to other endogenous selection mechanisms in estimating low pay transitions using the British Household Panel Survey. They find that simple models of endogenous selection produce similar estimates of covariate marginal effects than more complex models. On the other hand, Francesconi *et al.* (2009) show that attrition in the BHPS leads over time (eight years) to increasing divergence in results of typical earnings data analyses from those obtained using cross-sectional data.

possible to estimate the effect on transition estimates of time-invariant or other differences in unobserved characteristics between the sample used, attritors and those who do not have positive earnings in each wave.

Therefore, this paper takes a simpler approach. First, Table 1 compares the characteristics of the sample used for the main analyses with the characteristics of comparable people in the cross-sectionally weighted HLFS for the same years.

Formal tests of the statistical significance of differences have not been conducted. It appears, however, that compared to the cross-sectional data, women in the longitudinally weighted HLFS LIS are significantly more likely to be of European ethnicity, and men are more likely to be married. Other differences are unlikely to be or to be only marginally significant.

Second, kernel density plots of the logarithm of hourly earnings using the longitudinally weighted and cross-sectionally weighted data are compared (Figures 1a & 1b). Generally, the differences are small.

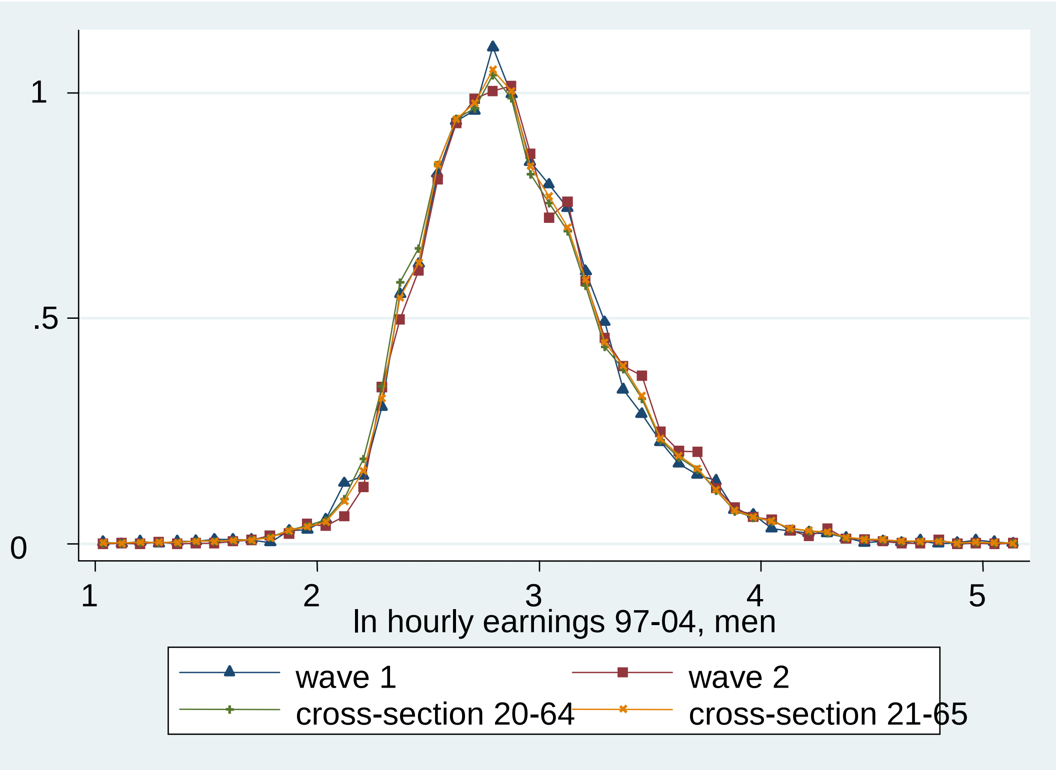
Previous studies and the comparisons presented here suggest that attrition and selection are unlikely to have a large effect on estimated transitions. Nevertheless, the regression analyses should best be interpreted as representing multivariate associations with earnings changes, and not necessarily causal influences.

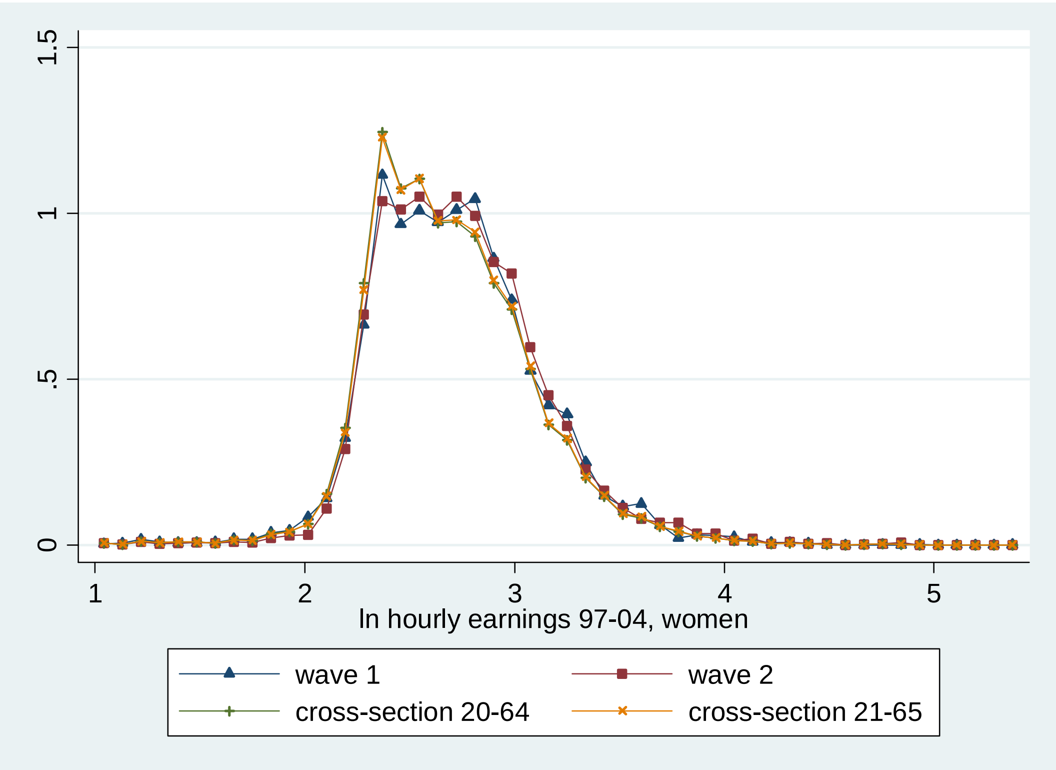
Table 1: Sample Characteristics: Individuals with Positive Hourly Earnings, Pooled Data 1997 - 2004

	Men		Women	
	Longitudinal	Cross-section	Longitudinal	Cross-section
Age	39.5	39.2	40.0	40.0
	(0.17)	(0.09)	(0.15)	(0.08)
20-24	0.09	0.11	0.09	0.10
	(0.005)	(0.003)	(0.004)	(0.002)
25-29	0.15	0.13	0.14	0.12
	(0.006)	(0.003)	(0.005)	(0.002)
30-39	0.28	0.29	0.24	0.26
	(0.006)	(0.003)	(0.006)	(0.003)
40-49	0.26	0.25	0.31	0.30
	(0.006)	(0.03)	(0.006)	(0.003)
50-64	0.22	0.22	0.22	0.23
	(0.006)	(0.003)	(0.005)	(0.003)
Maori	0.08	0.09	0.08	0.09
	(0.003)	(0.002)	(0.003)	(0.002)
Pacific	0.05	0.05	0.04	0.05
	(0.003)	(0.001)	(0.003)	(0.001)
European	0.81	0.80	0.83	0.81
	(0.005)	(0.003)	(0.005)	(0.003)
No qualifications	0.18	0.18	0.17	0.18
	(0.005)	(0.003)	(0.005)	(0.003)
School qualifications	0.21	0.20	0.27	0.26
	(0.006)	(0.003)	(0.006)	(0.003)
Post-school non-degree qualifications	0.46	0.46	0.42	0.41
	(0.007)	(0.004)	(0.007)	(0.004)
Degree	0.16	0.16	0.14	0.15
	(0.005)	(0.003)	(0.005)	(0.003)
Married	0.75	0.71	0.71	0.71
	(0.007)	(0.003)	(0.006)	(0.003)
Hourly Earnings	$20.10	$20.05	$16.58	$16.70
	(0.16)	(0.10)	(0.11)	(0.08)
Change in Hourly Earnings	$0.64	-	$0.71	-
	(0.11)		(0.14)	
Weekly Earnings	$858.32	$846.75	$553.85	$540.11
	(7.57)	(4.57)	(4.83)	(2.72)
Change in Weekly Earnings	$29.65	-	$22.74	-
	(4.01)		(2.96)	
Observations	7851	24973	8725	26799

Note: – pooled longitudinally weighted data for individuals aged 20-64 in the first year, with positive hourly earnings in each year and a complete set of the covariates used in the later regression analysis, are compared to pooled cross-sectionally weighted data for those aged 20-64 with positive earnings. Standard errors (in parentheses) have been estimated using Stata's `svytab` command. Monetary sums are adjusted for inflation to $1997. The cross-sectional data includes two observations each for those included in the longitudinal data, and also includes those who have positive hourly earnings in only one year of any two in the LIS.

Figure 1: **Comparison of Kernel Density Plots of Log Hourly Earnings for Cross-sectionally and Longitudinally Weighted Data - 1997 - 2004**





Annual transitions across quintiles in weekly earnings and hourly earnings

This section first compares annual transitions in LIS weekly earnings with annual earnings transitions reported in extensions to Hyslop (2000a) using tax data.[29] The main analysis is restricted to individuals aged 20-64 in the first year, who have positive earnings in each year – 10,375 men and 9,829 women. The purpose is to get a sense of how far the LIS can be relied on to measure annual transitions in hourly earnings, and whether differences between the LIS and tax data can potentially be accounted for by known differences in the datasets and the earnings measures used.

To ensure the analyses are as comparable as possible to the tax data results, inflation adjusted weekly earnings are first adjusted for year and age,[30] and then partitioned into five quintile groups in each wave of each of the seven sets of linked data. Annual transitions across quintile groups are estimated using pooled data covering the years 1997 – 2004, separately for men and women. A similar approach is then taken to estimate transition matrices for annual changes in LIS hourly earnings.

Also to ensure greater comparability with the tax data analyses, weekly earnings are defined as the sum of weekly self-employment income, and usual weekly earnings from all jobs excluding self-employment income. (Hourly earnings in subsequent analyses are defined as usual hourly earnings). Weekly and hourly earnings are, in all analyses, adjusted by the Consumer Price Index to 1997 dollars.

Comparisons are made with tax data of the probability of remaining in the original earnings quintile group between years for those with positive earnings in both years. The tax data is calculated from transitions reported in Hyslop (2002) covering the

[29] The LIS annual market income measure is grouped in $5,000 bands up to $100,000 and hence is less precise than a weekly earnings measure, particularly when inflation adjustments are to be made and data pooled. As a result, the current research uses the weekly measure, while recognising that this is likely to capture very short run intra-year fluctuations that are smoothed out by an annual measure. Hyslop and Yahanpath (2005) provide more analysis of the HLFS IS annual income measure and its relationship to an annualised weekly income measure.

[30] The adjustment follows Hyslop (2000a) by regressing CPI adjusted earnings on year dummies and age, separately for men and women, and then constructing quintile groups of earnings using the residuals. Early analyses used quantile regressions for this purpose, but the resulting transition estimates were not materially different to analyses that used OLS. All the transition matrices estimated in this study use the same procedure.

years 1994-1998.[31] The LIS weekly earnings measure will capture some short term volatility that is likely to "ironed out" on an annual measure. In particular, the self-employment income measure in the LIS is likely to be more volatile than self-employment income in tax data. On both these counts, it can be expected that transition probabilities in the LIS weekly earnings measure will be higher than in the tax data, particularly for men for whom self-employment income is a more important source of earnings than for women.

The relationship between the weekly and hourly earnings measures also needs to be considered. The weekly earnings measure includes self-employment income, for which Statistics New Zealand does not derive an hourly earnings measure. As a result, in the pooled longitudinally weighted data, of those with positive weekly earnings, only 74 percent of men and 88 percent of women have positive hourly earnings recorded.[32] When the sample for the main hourly earnings analysis is restricted to individuals for whom a full set of covariates used in the regression analysis is available, this amounts to 7,851 men and 8,725 women.

Unreported analyses show that when self-employed income is removed from the LIS weekly earnings measure, earnings volatility between years is reduced, as expected. On the other hand, because of the way it is constructed, the hourly earnings measure is likely more influenced by variations in hours of work than is the weekly measure.[33] *A priori* it is not clear, therefore, whether transition probabilities will be higher or lower in the hourly earnings compared with the weekly earnings analyses. In addition the effects of these influences are likely to vary by gender and location in the distribution of earnings.

[31] Hyslop (2002) is an unpublished extension to Hyslop (2000a) helpfully provided to the author in a personal communication.

[32] Consistent with this in analyses not reported here, it was found that men with positive hourly earnings in each year are a significantly younger sub-set of those with positive weekly earnings (including self-employed income). They are also less likely to be of "European" ethnicity, to be married, and they appear to have slightly lower weekly market incomes. The differences for women are small.

[33] See footnote 4.

The determinants of annual changes in weekly and hourly earnings

This part of the analysis uses multivariate regression analysis to identify individual characteristics that are significantly associated with changes in earnings between years, controlling for a wide range of covariates. To the extent that these characteristics are stable over time they may represent processes that have significant effects on life-time earning profiles. Because their life-time earning profiles differ markedly, separate models for men and women were estimated throughout.

A range of preliminary analyses were undertaken to determine a suitable functional form. It was found that logistic and OLS regressions of the probability of having a significant change in earnings between years (defined as a 10 percent increase or decrease) produced qualitatively similar results, both requiring a relatively arbitrary definition of significant change in earnings, and involved separate regressions for positive and negative earnings changes. As a result, OLS was chosen for the main results, using a continuous measure of the change in hourly earnings between years as the dependent variable. To investigate whether the same factors influence both decreases and increases in earnings, OLS is supplemented by quantile regressions, with estimates at the 25th, 50th and 75th percentiles.[34] These represented a moderate negative, a small positive and a moderate positive change in earnings respectively.

A further issue was a choice between the change in levels and the change in log hourly earnings as the dependent variable. Preliminary analyses produced qualitatively similar results for the change in levels and the change in log hourly earnings. Dropping the highest and lowest one percent of observations by earnings in the first year also produced qualitatively similar results for analyses using the change in log earnings as the dependent variable. As a result, analyses with change in log hourly earnings as the dependent variable have been used.

The OLS regressions were conducted by sequentially adding blocks of variables. First demographic variables were entered – age group (with 40-49 the omitted category), ethnicity (Maori, Pacific, other and "not specified" with European the omitted

---

[34] Stata `sqreg` command was used to estimate these regressions, allowing the reporting of boot-strapped standard errors.

category),[35] whether married, and qualifications in four categories (with post-school non-degree qualifications the omitted category). *A priori* it might be expected that differences in human capital may be associated with differences in earnings growth year on year. Better educated people have more opportunity for employment based education and training, including "learning on the job". This should be reflected in faster growth in earnings than those with lower education. The measures of education available in the HLFS are very broad, and the effects of differences in human capital may thus be captured by other demographic variables correlated with these unmeasured differences. Age may also influence earnings growth, for instance through increasing or reducing labour market attachment.

Next controls were entered for year, household type, industry, region and occupation, to help isolate the effect of the demographic and qualification variables *per se*. Finally, a control was entered for log initial hourly earnings. This was intended to control for measurement error and/or regression to the mean in the hourly earnings measure.[36] In particular those on lower initial hourly earnings might, other things being equal, have a larger percentage increase in hourly earnings over a year. If log initial earnings is not controlled, then this effect could, for instance, disguise the effect of other characteristics (such as belonging to Maori and Pacific ethnicity, or having low qualifications) that are over-represented amongst those with low hourly earnings.

Sub-period comparison – 1997-2001 and 2001- 2004

Transition matrices and the full OLS regressions for change in log hourly earnings were repeated for each of the sub-periods 1997-2001 and 2001-2004.[37] In the second period the sample population was, on average, a little older, a little less likely to be European and less likely to have no qualifications than in the first (see Table A1). In 1997 and 1998 the New Zealand economy experienced a brief recession, while the

[35] Statistics New Zealand prioritised ethnic variables were used whereby anybody who identifies themselves as Maori is classified as Maori, and anybody remaining who identifies themselves as Pacific is classified as Pacific.
[36] Creedy (1997) finds evidence for regression to the mean in his analysis of income mobility using New Zealand tax data.
[37] Each sub-period though covering four and three years of data respectively, contained a similar number of observations because of the accelerated rotation out of the HLFS panel in 1998/1999 and 1999/2000.

period from 1999 to 2004 was a period of continuing economic growth and growth in per capita income of two to three percent per annum. Evidence suggests that transitory volatility in earnings of men increases in a recession, [38] so some differences in transition probabilities for men between the two sub-periods might be expected as a result.

6 Results

Annual transitions across quintiles in weekly and hourly earnings 1997-2004

The main focus of this paper is on annual changes in hourly earnings. The purpose of this section is to link findings on transitions in hourly earnings, to previous research on transitions in annual market income using tax data (Hyslop, 2000a, 2002). As explained above, weekly earnings in the LIS are used to make the link.

Tables 2a and 2b show that about 78 percent of men and 57 percent of women aged 20-64 have positive weekly earnings in both years of the LIS.[39]

[38] See Haider (2001), Gottschalk and Moffitt (2008) and Shin and Solon (2008) all using the PSID.
[39] For men, 93 % of the 83% with positive earnings in the first year also had positive earnings in the second year. The corresponding figures for women are 88% and 65%.

Table 2a Probability of Annual Transitions across Weekly Earnings States HLFS LIS – 1997-2004 - Men

First Year Earnings State	Second Year Earnings State	
	Non-positive	Positive
	0.178	0.822
Non-positive	**0.726**	0.274
0.169	(0.012)	(0.012)
Positive	0.067	**0.933**
0.831	(0.003)	(0.003)

Observations: 13,814

Table 2b Probability of Annual Transitions across Weekly Earnings States HLFS LIS – 1997-2004 - Women

First Year Earnings State	Second Year Earnings State	
	Non-positive	Positive
	0.361	0.639
Non-positive	**0.802**	0.198
0.355	(0.007)	(0.007)
Positive	0.118	**0.882**
0.645	(0.004)	(0.004)

Observations: 17,699

Note: Includes individuals aged 20-64 in the first year with non-missing data on weekly earnings in each year. Weekly earnings includes usual weekly earnings from wage and salary jobs, and self-employed income. Standard errors (in parentheses) have been estimated using Stata's `svytab` command. Diagonals show the probability of remaining in the original state in the second year. The estimated proportion of observations with positive earnings in both years differs from the estimate using Stata's `tab` command. Using this command, the number of such observations for men is 10,375 as represented in Table 3a, and for women is 9,829 as represented in Table 3b.

Table 3a and 3b show annual transitions in weekly earnings conditional on having positive earnings in both years. To assist comparisons with the IRD data, self-employed income is included in this measure.

Table 3a: Weekly Earnings Annual Transition Probabilities

HLFS LIS – 1997-2004 - Men

	Second Year Earnings Quintile Group				
First Year Earnings Quintile Group	1	2	3	4	5
1	**0.634**	0.216	0.076	0.040	0.034
2	0.203	**0.469**	0.220	0.075	0.034
3	0.068	0.206	**0.458**	0.214	0.055
4	0.042	0.071	0.218	**0.532**	0.137
5 (top)	0.023	0.033	0.051	0.183	**0.709**

Observations: 10,375

Table 3b: Weekly Earnings Annual Transition Probabilities

HLFS LIS – 1997-2004 - Women

	Second Year Earnings Quintile Group				
First Year Earnings Quintile Group	1	2	3	4	5
1	**0.725**	0.165	0.066	0.023	0.021
2	0.207	**0.531**	0.175	0.057	0.031
3	0.066	0.202	**0.489**	0.202	0.041
4	0.022	0.075	0.199	**0.551**	0.154
5 (top)	0.017	0.031	0.060	0.164	**0.728**

Observations: 9,829

Note: Includes individuals aged 20-64 in the first year who have positive weekly earnings in each year. Diagonal elements represent probability of remaining in the original quintile group in the second year. Standard errors on diagonal elements (estimated using Stata's `svytab` command) range from 0.012 to 0.014. Earnings have been inflation adjusted, and quintile groups constructed from the residuals obtained by regressing earnings on age and year categories.

Table 4 compares the probability of remaining in the original quintile group in the LIS and in tax data.

Except in the bottom fifth, the probability of men remaining in the original quintile group appears to be significantly higher in the tax data than in the LIS, but for women, the difference between the LIS and the IRD measure is smaller. The LIS measure compared to the IRD measure is likely to reflect greater transitory variation in weekly earnings measured at particular points during the year, compared to the annual income measure in the IRD which "irons out" these transitory variations.[40] There may also be greater error in the measurement of self-employed income in the LIS compared to IRD data.[41] This might explain the greater difference in persistence between the two measures for men (for whom self-employed income is more important), compared to women.

Table 4: Probability of Remaining in Original Quintile Group of Earnings Conditional on having Positive Earnings in Each Year – LIS (1997 – 2004) compared to Tax Data (1994-1998)*

	Men			Women		
Quintile Group	LIS (Weekly)	IRD (Annual)	% difference	LIS (Weekly)	IRD (Annual)	% difference
1	0.634	0.622	-2	0.725	0.594	-18
2	0.469	0.525	+12	0.531	0.496	-7
3	0.458	0.556	+21	0.489	0.516	+6
4	0.532	0.609	+14	0.551	0.590	+7
5 (top)	0.709	0.766	+8	0.728	0.767	+5

* calculations based on Hyslop (2002).

Tables 5a, 5b and 6a and 6b show annual hourly earnings transitions. About 61 percent of men and 49 percent of women aged 20-64 have positive hourly earnings in both years.[42]

[40] The LIS weekly earnings measure used here also excludes some people who temporarily do not have positive weekly earnings, but who do have earnings on an annual basis. Hyslop and Yahanpath (2005) show that such people are over-represented in very low income groups on the annual earnings measure. The people in the bottom quintile group in the LIS weekly earnings measure used here are likely to be less comparable to the corresponding group on an annual earnings measure, than in other parts of the distribution.

[41] Unreported analyses show that, consistent with this, when those with only self-employed income are excluded from the LIS analysis, the probability of remaining in the original quintile group increases.

[42] For men, 89 % of the 62% with positive hourly earnings in the first year also had positive earnings in the second year. The corresponding figures for women are 87% and 57%.

Table 5a Probability of Annual Transitions across Hourly Earning States

HLFS LIS – (1997-2004) - Men

First Year Earnings State	Second Year Earnings State	
	Zero	Positive
	0.400	0.600
Zero	**0.873**	0.127
0.383	(0.006)	(0.006)
Positive	0.107	**0.893**
0.618	(0.004)	(0.004)

Observations: 13,758

Table 5b Probability of Annual Transitions across Hourly Earning States

HLFS LIS – (1997-2004) - women

First Year Earnings State	Second Year Earnings State	
	Zero	Positive
	0.443	0.557
Zero	**0.849**	0.151
0.433	(0.005)	(0.005)
Positive	0.132	**0.868**
0.567	(0.004)	(0.004)

Observations: 17,692

Note: Includes individuals aged 20-64 in the first year with non-missing data on hourly earnings in each year. Standard errors (in parentheses) estimated using Stata's `svytab` command. Diagonals show the probability of remaining in the original state in the second year. The implied numbers with positive hourly earnings in each year differs from those represented in Table 6a & b for two reasons: some observations are dropped in Table 6b and subsequent analyses because of missing variables used in the regression analyses; Stata's `svytab` command also weights the data differently to the `tab` command. When only observations with non-missing variables are included the `tab` command represents the actual number of observations with positive hourly earnings in each year (men: 7851; women: 8,725) that are used in subsequent analyses.

Table 6a Hourly Earnings Annual Transition Probabilities

HLFS LIS 1997-2004 - Men

First Year Earnings Quintile Group	Second Year Earnings Quintile Group 1	2	3	4	5
1	**0.708**	0.187	0.048	0.037	0.021
2	0.185	**0.489**	0.220	0.075	0.031
3	0.075	0.212	**0.455**	0.201	0.056
4	0.033	0.077	0.246	**0.495**	0.149
5 (top)	0.022	0.033	0.065	0.180	**0.700**

Observations: 7,851

Table 6b Hourly Earnings Annual Transition Probabilities

HLFS LIS 1997-2004 - Women

First Year Earnings Quintile Group	Second Year Earnings Quintile Group 1	2	3	4	5
1	**0.683**	0.228	0.049	0.027	0.013
2	0.223	**0.478**	0.202	0.064	0.033
3	0.063	0.189	**0.459**	0.221	0.069
4	0.036	0.061	0.212	**0.515**	0.176
5 (top)	0.022	0.027	0.083	0.217	**0.652**

Observations: 8,725

Note: Includes individuals aged 20-64 in the first year with positive hourly earnings in each year. Diagonal elements represent probability of remaining in the original quintile group in the second year. Standard errors on diagonal elements (estimated using Stata's `svytab` command) range from 0.013 to 0.017. Earnings have been inflation adjusted, and quintile groups constructed from the residuals obtained by regressing earnings on age and year categories.

Table 7 compares the probability of remaining in the original quintile group in the weekly and hourly earnings data. For men, the differences are not significant, except in the bottom quintile group where the probability of remaining is higher for hourly earnings. For women, the probability for remaining in the original quintile group is systematically lower for hourly earnings than for weekly earnings, across all quintile groups.

It is likely that, at least for men, the differences can be accounted for by two balancing influences. First, the removal of self-employed in come in the hourly measure is likely to reduce volatility. On the other hand, the hourly measure is more influenced by variations in hours of work (as almost two thirds of individuals have hourly earnings calculated as usual weekly earnings divided by hours of work).[43] Hours worked are likely to be more volatile than directly measured hourly wages or weekly salaries.[44]

Table 7: Probability of remaining in Original Quintile Group Conditional on having Positive Earnings in Each Year – Weekly Compared to Hourly Earnings 1997-2004

	Men			Women		
Quintile Group	Weekly Earnings	Hourly Earnings	% difference	Weekly Earnings	Hourly Earnings	% difference
1	0.634	0.708	+12	0.725	0.683	-6
2	0.469	0.489	+4	0.531	0.478	-10
3	0.458	0.455	-1	0.489	0.459	-6
4	0.532	0.495	-7	0.551	0.515	-7
5 (top)	0.709	0.700	-1	0.728	0.652	-10

In sum, (except for the bottom quintile group) men are more likely to move out of their original quintile group in the LIS weekly earnings data, than in the IRD annual earnings data.[45] This might be expected, as the annual earnings measure in the IRD 'irons out' weekly fluctuations in earnings that probably remain in the LIS weekly earnings measure.[46] Except for the bottom quintile group, LIS hourly earnings transitions for men are similar in magnitude to LIS weekly earnings transitions. This is likely to be accounted for by the fact that the hourly earnings measure does not include self-employed income, but is likely to be more influenced by variations in hours of work than the weekly measure.

[43] See footnote 4.

[44] Haider (2001), using data on men in the PSID, shows that changes in earnings stability in the United States in the 1970s are mostly accounted for by changes in volatility in hours of work, rather than in hourly wages. Consistent with this, unreported sub-period analyses using the weekly earnings measure show no difference in transition probabilities between 1997-2001 (which includes the 1997-1998 recession) and 2001-2004. When an hourly earnings measure is used there are differences for men, as the results reported later in this section show.

[45] The difference in the bottom quintile group might be because those with temporarily zero weekly earnings are excluded from the LIS data, but may be included in the IRD data. (see Hyslop and Yahapanth, 2005).

[46] When self-employed income is removed from the LIS weekly earnings measure, the probability of remaining in the original quintile group rises, suggesting that measurement error or transitory variation in self-employment income could explain some of the differences between the IRD and LIS data. Self-employment income is a much smaller proportion of earnings for women.

While, for women, there is little difference between the IRD annual earnings transitions and the LIS weekly earnings transitions, the probability of remaining in the original quintile group is always lower on the hourly earnings measure.

Annual variations in hourly earnings are likely to reflect a range of factors. These include age (and hence the stage in the working life, with changes likely to be more marked at the beginning and end of working life), qualifications (and hence scope for advancement), other dimensions of human capital that might not be captured in qualifications measures, marital status and household type (and hence willingness and opportunity to vary hours of work, and to seek higher paid work) occupation, industry and region (reflecting prevailing conditions in labour and product markets) and whether the individual is in an occupation where earnings are paid as a wage or a salary (and hence whether measured hourly earnings are influenced by reported hours of work). The next section reports the results of multivariate analyses of these influences.

The determinants of annual changes in weekly and hourly earnings

Tables 8a and 8b show regression results for men and women respectively for the determinants of change in log hourly earnings, for those with positive hourly earnings in each year. The first three columns show OLS results, with controls for household type, labour market, and log initial earnings being added successively.[47] The last three columns look for differences in effects at different quantiles of the change in log hourly earnings.

For men, having a degree always has a positive effect on earnings transitions, which becomes quite strong once initial earnings are controlled.[48] With this control, being married also has a positive effect, while having no or school qualifications, and being

[47] In earlier research, household type was added in a separate step to labour market variables, but there were no consistent and clear results, so this step has been omitted.

[48] The implied earnings advantage after one year for men with a degree (compared with men with another post-school qualification) with mean hourly earnings in the first year is 11%, and, for women, 8%. Many more periods of data, and a more sophisticated model that accounts for regression to the mean and serial correlation in the evolution of earnings would be required to estimate the long term earnings variations associated with different educational qualifications.

a Pacific person has a negative effect. The co-efficient on log hourly earnings is always significant and negative. Without this control, the effect of being a Pacific person and having low qualifications (both of which are associated with lower earnings) is not significant. The negative effect of being a Pacific person appears to be stronger at the 25th percentile of change (a drop in earnings) than at other points in the distribution, while the effects of no qualifications and of being married appear to be strongest at the 75th percentile.

For women, having a degree only has a significant positive effect on earnings transitions, and having no qualifications only has a significant negative effect, when initial earnings are controlled. The effects are similar across the distribution of change in log hourly earnings, with some suggestion that low qualifications have a greater effect at the 75th percentile of change than at the 25th percentile or median. Having school qualifications has a significant negative effect only in the quantile regressions at the mean and 75th percentile. Being a Pacific person only has a significant negative effect at the mean/median when initial earnings are controlled. In contrast to the results for men, being aged over 50 always has a negative effect for women,[49] while being married only has a significant positive effect in the quantile regressions at the median.

In all regressions demographic and qualifications variables explained very little of the variation in the change in log hourly earnings. Adding household type and labour market variables increased the explanatory power modestly, with a much more substantial increase being accounted for by controlling for initial earnings. Measurement error is likely to account for a substantial part of the effect of log initial earnings on the change in log earnings.

[49] Earlier results also showed a negative effect for men, when the change in log weekly earnings (including self-employed income) was the dependent variable, suggesting that self-employment represents a reduced attachment to the labour market for many older working age men.

Table 8a: Determinants of Change in Log Hourly Earnings between Years, OLS and Quantile regressions, Men 1997-2004

	OLS (1)	OLS (2)	OLS(3)	Q25	Q50	Q75
Age 20-24	0.041	0.040	-0.056	-0.017	-0.004	-0.011
	(0.020)**	(0.021)*	(0.019)***	(0.014)	(0.007)	(0.012)
Age 25-29	0.051	0.046	-0.018	-0.010	-0.001	-0.010
	(0.014)***	(0.015)***	(0.014)	(0.010)	(0.008)	(0.009)
Age 30-39	0.029	0.027	0.009	0.002	0.002	0.003
	(0.010)***	(0.010)**	(0.009)	(0.006)	(0.006)	(0.005)
Age 50-64	0.015	0.011	-0.001	-0.009	-0.007	-0.016
	(0.013)	(0.013)	(0.011)	(0.007)	(0.005)	(0.009)*
Maori	-0.013	-0.008	-0.020	-0.012	-0.004	-0.006
	(0.012)	(0.012)	(0.011)*	(0.012)	(0.009)	(0.009)
Pacific	0.002	-0.006	-0.081	-0.068	-0.036	-0.045
	(0.016)	(0.018)	(0.016)***	(0.017)***	(0.013)***	(0.014)**
Other	0.052	0.044	-0.033	-0.037	-0.007	-0.010
	(0.019)***	(0.019)**	(0.019)*	(0.016)**	(0.010)	(0.022)
Unspecified	0.003	0.001	-0.032	0.036	0.002	-0.050
	(0.054)	(0.058)	(0.045)	(0.058)	(0.037)	(0.047)
Married	0.002	0.023	0.062	0.019	0.025	0.044
	(0.010)	(0.019)	(0.016)***	(0.014)	(0.008)**	(0.013)***
No Qualifications	-0.020	-0.014	-0.060	-0.029	-0.027	-0.050
	(0.011)*	(0.012)	(0.011)***	(0.008)***	(0.007)***	(0.009)***
School Quals	-0.006	-0.007	-0.021	-0.012	-0.010	-0.019
	(0.011)	(0.012)	(0.010)**	(0.008)	(0.006)*	(0.009)**
Degree	0.033	0.031	0.105	0.077	0.072	0.079
	(0.014)**	(0.015)**	(0.015)***	(0.012)***	(0.007)***	(0.012)***
Year Dummies	No	Yes	Yes	Yes	Yes	Yes
Household Type	No	Yes	Yes	Yes	Yes	Yes
Industry	No	Yes	Yes	Yes	Yes	Yes
Region	No	Yes	Yes	Yes	Yes	Yes
Occupation	No	Yes	Yes	Yes	Yes	Yes
Log Hrly Earnings	No	No	-0.421	-0.301	-0.223	-0.284
			(0.029)***	(0.018)***	(0.014)***	(0.014)***
Constant	0.006	-0.003	1.221	0.775	0.657	0.964
	(0.012)	(0.033)	(0.093)***	(0.052)***	(0.045)***	(0.046)***
Observations	7851	7851	7851	7851	7851	7851
(Pseudo) R squared	0.009	0.026	0.259	0.088	0.060	0.087

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: Individuals aged 20-64 in the first year, with positive hourly earnings in each year. OLS results estimated using Stata's `svyreg` command. Quantile regressions estimated using Stata's `sqreg` command. The change in log hourly earnings were 0.066. 0.019 and 0.134 at 25th, 50th and 75th percentiles respectively. Omitted categories are Age 40-49, European, Unmarried and Post-school non-degree qualifications. The dependent variable represents the natural logarithm of the ratio of second year to first year hourly earnings. The coefficients (βs) on categorical variables may be interpreted in terms of the percentage change in this ratio due to belonging to the particular category compared to belonging to the omitted category, according to the formula (exp(β)-1)*100.

Table 8b: Determinants of Changes in log Hourly Earnings between Years, OLS and Quantile regressions, Women 1997-2004

	OLS (1)	OLS (2)	OLS(3)	Q25	Q50	Q75
Age 20-24	0.014	0.012	-0.060	-0.022	-0.012	-0.030
	(0.021)	-0.023	(0.022)***	(0.010)**	(0.010)	(0.012)***
Age 25-29	0.004	0.006	-0.016	-0.009	-0.003	-0.012
	(0.014)	(0.015)	(0.014)	(0.008)	(0.005)	(0.009)*
Age 30-39	-0.014	-0.009	-0.003	0.004	-0.001	0.005
	(0.011)	(0.011)	(0.010)	(0.005)	(0.004)	(0.007)
Age 50-64	-0.038	-0.041	-0.043	-0.021	-0.022	-0.024
	(0.010)***	(0.011)***	(0.010)***	(0.006)***	(0.005)***	(0.007)***
Maori	0.017	0.017	-0.007	-0.004	0.000	0.020
	(0.014)	(0.015)	(0.013)	(0.009)	(0.006)	(0.009)*
Pacific	-0.004	-0.002	-0.029	-0.037	-0.022	-0.021
	(0.017)	(0.018)	(0.017)*	(0.024)	(0.010)*	(0.016)
Other	-0.001	-0.001	-0.023	-0.015	0.000	0.007
	(0.028)	(0.028)	(0.028)	(0.012)	(0.011)	(0.014)
Unspecified	0.035	0.025	0.057	-0.014	0.002	0.123
	(0.053)	(0.056)	(0.074)	(0.108)	(0.109)	(0.170)
Married	0.000	0.011	0.025	0.012	0.016	0.016
	(0.010)	(0.016)	(0.016)	(0.010)	(0.008)**	(0.013)
No Qualifications	-0.004	-0.005	-0.052	-0.024	-0.026	-0.052
	(0.010)	(0.011)	(0.011)***	(0.007)***	(0.004)***	(0.006)***
School Quals	0.009	0.004	-0.006	-0.010	-0.013	-0.017
	(0.010)	(0.011)	(0.009)	(0.007)	(0.004)***	(0.007)**
Degree	0.006	0.006	0.079	0.069	0.058	0.088
	(0.015)	(0.017)	(0.016)***	(0.011)***	(0.008)***	(0.011)***
Year Dummies	No	Yes	Yes	Yes	Yes	Yes
Household Type	No	Yes	Yes	Yes	Yes	Yes
Industry	No	Yes	Yes	Yes	Yes	Yes
Region	No	Yes	Yes	Yes	Yes	Yes
Occupation	No	Yes	Yes	Yes	Yes	Yes
Log Hrly Earnings	No	No	-0.432	-0.379	-0.253	-0.342
			(0.022)***	(0.028)***	(0.019)***	(0.017)***
Constant	0.036	0.034	1.263	1.004	0.744	1.102
	(0.012)***	(0.029)	(0.066)***	(0.081)***	(0.054)***	(0.048)***
Observations	8725	8725	8725	8725	8725	8725
(Pseudo) R squared	0.004	0.013	0.191	0.096	0.062	0.100

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: Individuals aged 20-64 in the first year, with positive hourly earnings in each year. OLS results estimated using Stata's `svyreg` command. Quantile regressions estimated using Stata's `sqreg` command. The change in log hourly earnings was 0.059. 0.012 and 0.117 at 25th, 50th and 75th percentiles respectively. Omitted categories are Age 40-49, European, Unmarried and Post-school non-degree qualifications. The dependent variable represents the natural logarithm of the ratio of second year to first year hourly earnings. The coefficients (βs) on categorical variables may be interpreted in terms of the percentage change in this ratio due to belonging to the particular category compared to belonging to the omitted category, according to the formula (exp(β)-1)*100.

Sub-period comparison – 1997-2001 and 2001- 2004

For both men and women, a significantly higher proportion of the sample had positive hourly earnings in both years in the later sub-period compared to the first. For men, the proportion rose from 53 percent to 58 percent, and for women, from 48 percent to 52 percent. The sample individuals were on average older in the second sub-period and less likely to be European. They were also more likely to have post-school qualifications. Weekly and hourly earnings were substantially higher (Table A1).

Men in the upper part of the distribution appear to have a higher probability of remaining in their original quintile group in the second period compared to the first. Any differences for women are small and not significant (Appendix Table A2).[50] Overall, this suggests, at least for women, and for men in the lower half of the distribution of hourly earnings, stability in transition probabilities over time. The increased probability of remaining in the same quintile group in the upper part of the distribution for men might be due to cyclical effects. The second period was one of sustained economic growth in New Zealand, and evidence from the United States (cited previously) suggests that transitory earnings volatility increases in a recession.

The OLS regression results were qualitatively similar for sub-periods for both men and women. For men, though, having a school qualification only had a significant negative effect in the second sub-period, due to a large change in the size of the co-efficient (see Appendix Table A3). This may be related to changes in the composition of qualifications held by the sub-sample, and thus in the relationship of these to unmeasured learning ability, and thus to the rate of earnings growth.

7 Discussion and conclusion

This paper set out to investigate the annual transitions in hourly earnings using the LIS. In particular, it looks at movements up and down the earnings distribution and the demographic and human capital characteristics associated with those movements. A first issue that needs to be considered is the possible effect of attrition and sample selection on the estimates.

[50] Earlier analyses looked at transitions in weekly earnings and found no significant differences between sub-periods for either men or women. This suggests that the sub-period differences for hourly earnings transitions may be due to the greater influence of volatility in hours of work on the hourly earnings measure (see footnote 4).

Attrition and sample selection bias

The nature of the HLFS and the research strategy which focuses on individuals with positive hourly earnings in each year of the LIS leads to substantial sample selection. Other work cited above suggests that attrition over one year is likely to have a less important effect on estimates than other sources of sample selection. In any case, given only two periods of data, these issues cannot be credibly addressed econometrically. However, comparison of the observable characteristics of the sample used in the main analyses with similar individuals from the cross-sectionally weighted HLFS data show only small differences. Similarly, there are only small differences in the distribution of income.

Comparison of transition probabilities – LIS versus IRD data

The differences in the estimated probability of remaining in the original earnings quintile group between the LIS and IRD appear to be able to be explained by a range of factors. The weekly earnings measure in the LIS captures transitory variation that is smoothed out in the annual earnings measure used in the tax data (Hyslop, 2000, 2002). Self-employment income is also arguably likely to be measured with less error in the tax data than in the LIS weekly earnings measure. Comparisons across the bottom quintile group in the two sets of analysis show a different relationship which may reflect the fact that people with temporarily zero weekly earnings (but positive annually earnings) are missing from the LIS analysis. These people tend to be over-represented among low income earners (Hyslop and Yahanpath, 2005).

The differences in LIS weekly and hourly earnings transition probabilities are likely to be able to be explained by two factors with opposite effects on volatility that vary by gender, and by location in the distribution of earnings. The weekly earnings measure includes self-employed income, and unreported analyses show that including self-employment income increases the volatility of annual transitions in weekly earnings. This influence is more important for men than women. On the other hand, almost two-thirds of observations of the hourly earnings measure are calculated from usual weekly earnings and usual hours of work.[51] As a result, the weekly earnings measure may be less volatile between years than the hourly measure, for those receiving a salary and reporting varying hours of work.

Overall, allowing for these sources of differences, the LIS data shows a similarly moderate degree of year-on-year earnings mobility as the tax data.

The determinants of changes in earnings

Because there are only two periods of data, it is not possible to produce credible estimates of the causes of annual changes in hourly earnings. A range of multivariate analyses, however, produce reasonably stable estimates of the factors significantly associated with changes in hourly earnings. Together, however, these explain only a small proportion of the variation in changes in earnings. Measurement error in hourly earnings is likely to explain a large part of the variation, and be reflected in the estimated significant negative effect of initial hourly earnings on the change.

The most consistent and stable result is the positive effect of having a degree on change in log hourly earnings from one year to the next (compared to having a post-school non degree qualification). For men, this was significant in all regressions, and for women, in all regressions where log initial hourly earnings was controlled. For both men and women, having no qualification had a consistently negative effect, significant in almost all regressions for men, and for women in regressions where log initial hourly earnings was controlled. When log intial hourly earnings is controlled, school qualifications also had a weak negative effect, significantly so in some regressions.

These results are consistent with the literature that find heterogeneous earnings growth profiles that, amongst other things, differ by early educational qualifications attained. This literature suggests that the annual earnings growth advantage associated with higher educational qualifications is persistent over time.

[51] See footnote 4 above.

Once log initial hourly earnings were controlled, being of Pacific ethnicity had a highly significant negative effect for men, but only weakly so and rarely significant for women. Pacific people are younger, have on average lower qualifications and thus lower average hourly earnings than the general population, so without controls for initial hourly earnings, the effect of measurement error or regression to the mean dominates the effect associated with ethnicity.

Being married had a significant positive effect for men, but not usually for women. Being aged 50-64 always had a strong and significant negative effect for women, but not for men, except in one regression that controlled for log initial hourly earnings.

These results suggest that human capital proxied by qualifications has an effect on earnings growth. People with higher qualifications are likely to have greater learning ability and to be given better opportunities for employment-based learning, and to thus have faster earnings growth with experience. Ethnicity may capture differences in aspects of human capital (such as English language skills, fields of study, level of study attained, and grades achieved) not measured well by broadly defined qualifications. For men, the positive effect of being married could reflect either greater attachment to the labour market, or unmeasured aspects of human capital. The effect of aging for women is likely to reflect reducing attachment to the labour market.

Sub-period analyses – 1997-2001 and 2001-2004

For men, the probability of remaining in the original quintile group of earnings was lower in the first sub-period, possibly reflecting the effects of the 1997/98 recession. Studies based on United States data for men show that earnings volatility is counter-cyclical. The differences for women were small and not significant. The regression results were broadly consistent across sub-periods for both men and women.

Conclusion

This research shows that LIS estimates of annual weekly earnings transitions for those with positive weekly earnings in each year are reasonably consistent with estimates based on tax data, especially for women. As might be expected, there appears to be greater transitory volatility in the LIS measure of weekly earnings, which needs to be taken into account in interpreting results. LIS hourly earnings estimates of the probability of remaining in the original quintile group do not differ systematically from those based on weekly earnings. Any differences can potentially be explained by the inclusion of self-employment income in the weekly earnings measure but not the hourly earnings measure, and by the composite nature of the hourly earnings measure. The first increases the one-year volatility of weekly earnings measure relative to the hourly earnings measure, while the second is likely to increase the volatility of the hourly earnings measure relative to the weekly. The balance between these influences is likely to vary by gender and by place in the distribution of earnings.

Compared to tax data, the LIS offers a wide range of demographic, household and labour market variables with which to investigate factors associated with changes in hourly earnings. The results are consistent with differences in human capital and labour market attachment explaining a small proportion of the variation in annual changes in hourly earnings. Consistent with the literature, the differences in annual earnings growth associated with human capital differences could account for substantial differences in life-time earnings. Nevertheless, the fact that there are only two periods of data severely limits the extent to which causal relationships can be estimated and, obviously, the extent to which longer term earnings growth patterns can be determined.

Now that four years of data from SoFIE is available it should be possible to use more sophisticated estimation methods to untangle causality and longer term patterns of earnings growth. In contrast to SoFIE however, the LIS offers a large and ongoing source of longitudinal data that can reasonably credibly be used to estimate the magnitudes of annual earnings transitions, if not their causes.

References

Altonji, Joseph G., Anthony Smith and Ivan Vidangos (2009), "Modeling earnings dynamics", NBER Working Paper 1473, Cambridge MA: National Bureau of Economic Research, http://www.nber.org/papers/w14743 .

Atkinson, Anthony B., François J. Bourguignon and Christian Morrisson (1992), *Empirical studies of earnings mobility*. Philadelphia: Harwood.

Baker, Michael (1997), "Growth rate heterogeneity and the covariance structure of life-cycle earnings", *Journal of Labor Economics* 15(2): 537-579.

Baker, Michael and Gary Solon (2003), "Earnings dynamics and inequality among Canadian men, 1976-1992: Evidence from longitudinal income tax records", *Journal of Labor Economics* 21(2): 289-321.

Ballantyne, Suzie, Simon Chapple, David C. Maré and Jason Timmins (2003), "Movements into and out of child poverty in New Zealand: Results from the linked income supplement", Motu Working Paper 03-13, Wellington: Motu Economic and Public Policy Research, http://www.motu.org.nz/publications/detail/movements_into_and_out_of_child_poverty_in_nz .

Böheim, Renė and Stephen P. Jenkins (2006), "A comparison of current and annual measures of income in the British Household Panel Survey", *Journal of Official Statistics* 22(4): 733-758.

Bradbury, Bruce, Stephen P. Jenkins and John Micklewright (2001), "The dynamics of child poverty in seven industrialised nations", Chapter 4 in Bradbury, Bruce, Stephen P. Jenkins and John Micklewright *The dynamics of child poverty in industrialised countries*, Cambridge: Cambridge University Press: 92-132.

Burkhauser, Richard V., Douglas Holtz-Eakin and Stephen E. Rhody (1997), "Labor earnings mobility and inequality in the United States and Germany during the growth years of the 1980s", *International Economic Review* 38(4): 775-794.

Capellari, Lorenzo and Stephen P. Jenkins (2008), "Estimating low pay transition probabilities accounting for endogenous selection mechanisms", *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57(2): 165-186.

Creedy, John (1997), *Statics and dynamics of income distribution in New Zealand*, Wellington, Institute of Policy Studies.

Fitzgerald, John, Peter Gottschalk and Robert Moffitt (1998), “The impact of attrition in the panel study of income dynamics on intergenerational analysis”, *Journal of Human Resources* 33(2): 300-344.

Francesconi, Marco, Holly Sutherland and Francesca Zantomio (2009), “A comparison of earnings measures from longitudinal and cross-sectional surveys: Evidence from the UK”, ISER Working Paper 2009-14, Colchester: Institute of Social & Economic Research, http://www.iser.essex.ac.uk/publications/working-papers/iser/2009-14 .

Geweke, John and Michael Keane (2000), “An empirical analysis of earnings dynamics among men in the PSID: 1968-1989”, *Journal of Econometrics* 96(2): 293-356.

Gottschalk, Peter and Robert Moffit, (2008). “Trends in the transitory variance of male earnings in the U.S., 1970-2004”, Boston College Working Papers in Economics 697, Newton, MA.: Boston College, http://fmwww.bc.edu/EC-P/WP697.pdf.

Guvenen, Fatih, (2007), “Learning your earning: are labor income shocks really very persistent?”, *American Economic Review* 97(3): 687-712.

Haider, Steven J., (2001) “Earnings instability and earnings inequality of males in the United States: 1967-1999”, *Journal of Labor Economics* 19(4): 799-836.

Hugget, Mark, Gustavo Ventura and Amir Yaron (2002), “Human capital and earnings distribution dynamics”, NBER Working Paper 9366, Cambridge MA.: National Bureau of Economic Research, http://www.nber.org/papers/w9366.

Hugget, Mark, Gustavo Ventura and Amir Yaron (2007), “Sources of lifetime inequality”, NBER Working Paper 13224, Cambridge MA,: National Bureau of Economic Research, http://www.nber.org/papers/w13224.

Hyslop, Dean (2000a), "A preliminary analysis of the dynamics of individual market and disposable incomes", Treasury Working Paper 00/15, Wellington: New Zealand Treasury, <http://www.treasury.govt.nz/workingpapers/2000/00-15.asp>.

Hyslop, Dean (2000b) "Does benefit receipt affect future income? An econometric explanation", Treasury Working Paper 00/14, Wellington: New Zealand Treasury, <http://www.treasury.govt.nz/workingpapers/2000/00-14.asp>.

Hyslop, Dean (2002), "Update of Table 4A of 'A preliminary analysis of the dynamics of individual market and disposable incomes' to include data for 1998", unpublished personal communication.

Hyslop, Dean and Suresh Yahanpath (2005), "Income growth and earnings variations in New Zealand, 1998-2004", Wellington: New Zealand Treasury, working Paper 05/11, <http://www.treasury.govt.nz/workingpapers/2005/05-11.asp>.

Jenkins, Stephen and Christian Schluter (2003), "Why are child poverty rates higher in Britain than in Germany? A longitudinal perspective", *Journal of Human Resources* 38(2): 441-465.

Maani, Sholeh A. (2000), "Secondary and tertiary educational attainment and income levels for Maori and Non-Maori over time," Treasury Working Paper 00/18, Wellington: New Zealand Treasury, <http://www.treasury.govt.nz/workingpapers/2000/00-18.asp>

Maloney, Tim (2001), "Revised final report on family income dynamics in the Christchurch Health and Development Study", Unpublished report, Wellington: New Zealand Treasury.

Maloney, Tim and George Barker (1999), "The low-income dynamics of families in the Christchurch Health and Development Study", Unpublished report, Wellington: New Zealand Treasury.

O'Dea, Des (2000), "The changes in New Zealand's income distribution", Treasury Working Paper 00/13, Wellington: New Zealand Treasury, <http://www.treasury.govt.nz/workingpapers/2000/00-13.asp>.

Shin, Donggyun and Gary Solon (2008), "Trends in men's earning volatility: What does the Panel Study of Income Dynamics Show?", NBER Working Paper 14075, Cambridge MA.: National Bureau of Economic Research, http://www.nber.org/papers/w14075.

Smith, Harry and Robert Templeton (1990), "A longitudinal study of incomes", Unpublished report, Wellington: Department of Statistics.

Solon, Gary (1999), "Intergenerational mobility in the labor market", in Orley C. Ashenfelter and David Card (eds.) *Handbook of Labor Economics* Vol. 3A, Amsterdam: North Holland: 1761-1800.

Solon, Gary, Robert Barsky and Jonathan A. Parker (1994) "Measuring the cyclicality of real wages: How important is composition bias?", *Quarterly Journal of Economics* 109(1): 1-25.

Appendix: Sub-period comparisons – 1997-2001 versus 2001-2004

Table A1 : Sample Characteristics for Sub-periods 1997 – 2000 and 2001 - 2004

		Men		Women	
		1997-2000	2001-2004	1997-2000	2001-2004
Age		39.1	40.0	39.4	40.6
		(0.26)	(0.21)	(0.23)	(0.20)
		0.09	0.09	0.09	0.08
	20-24	(0.008)	(0.006)	(0.007)	(0.006)
		0.16	0.14	0.16	0.12
	25-29	(0.009)	(0.007)	(0.009)	(0.006)
		0.28	0.27	0.24	0.25
	30-39	(0.009)	(0.008)	(0.008)	(0.007)
		0.25	0.27	0.31	0.32
	40-49	(0.009)	(0.008)	(0.009)	(0.008)
		0.22	0.23	0.20	0.23
	50-64	(0.008)	(0.007)	(0.008)	(0.007)
Maori		0.07	0.09	0.08	0.09
		(0.005)	(0.005)	(0.005)	(0.004)
Pacific		0.04	0.06	0.04	0.05
		(0.005)	(0.004)	(0.004)	(0.004)
European		0.84	0.79	0.84	0.81
		(0.008)	(0.007)	(0.007)	(0.007)
No qualifications		0.18	0.17	0.19	0.16
		(0.008)	(0.006)	(0.007)	(0.006)
School qualifications		0.21	0.21	0.28	0.26
		(0.009)	(0.007)	(0.009)	(0.008)
Post-school non-degree		0.46	0.47	0.41	0.43
qualifications		(0.011)	(0.009)	(0.010)	(0.008)
Degree		0.15	0.16	0.13	0.16
		(0.008)	(0.007)	(0.007)	(0.006)
Married		0.76	0.74	0.72	0.70
		(0.010)	(0.008)	(0.009)	(0.008)
Hourly Earnings		$19.66	$20.60	$16.17	$17.05
		(0.23)	(0.23)	(0.15)	(0.16)
Change in Hourly Earnings		$0.69	$0.59	$0.63	$0.17
		(0.16)	(0.15)	(0.18)	(0.16)
Weekly Earnings		$840.80	$878.64	$536.42	$573.92
		(10.84)	(10.44)	(6.91)	(6.65)
Change in Weekly Earnings		$30.65	$28.51	$26.02	$19.19
		(5.64)	(5.68)	(3.71)	(4.73)
Observations		3577	4274	3964	4762

Note: – pooled longitudinally weighted linked data for individuals aged 20-64 in first year and with positive hourly earnings in each year, standard errors in parentheses, estimated using Stata's `svytab` command. Monetary sums are adjusted for inflation to $1997.

Table A2: Probability of Remaining in Original Hourly Earnings Quintile Group Conditional on having Positive Hourly Earnings in Each Year – 1997-2001 Compared to 2001-2004

	Men			Women		
Quintile Group	1997-2001	2001-2004	% difference	1997-2001	2001-2004	% difference
1	0.678	0.681	0	0.661	0.649	-2
2	0.459	0.452	-2	0.459	0.464	+1
3	0.414	0.431	+4	0.463	0.440	-5
4	0.452	0.515	+14	0.511	0.487	-5
5 (top)	0.643	0.726	+13	0.676	0.645	-5
Observations	3,577	4,274		3,964	4,762	

Note: Includes individuals aged 20-64 in the first year with positive hourly earnings in each year. Standard errors, estimated using Stata's `svytab` command vary from 0.017 – 0.026. Earnings have been inflation adjusted, and quintile groups constructed from the residuals obtained by regressing earnings on age and year categories. Between 1997-2001 and 2001-2004, the percentage of men aged 20-64 who had positive hourly earnings in each of the two years increased from 53% to 58%; the percentage of women aged 20-64 who had positive hourly earnings in each of the two years increased from 48% to 52%.

Table A3: Determinants of Changes in Log Hourly Earnings Between Years, OLS: 1997-2001 and 2001-2004

	Men		Women	
	97-01	01-04	97-01	01-04
Age 20-24	-0.078	-0.038	-0.050	-0.072
	(0.028)***	(0.024)	(0.019)***	(0.043)*
Age 25-29	-0.018	-0.020	-0.032	-0.002
	(0.020)	(0.017)	(0.017)*	(0.023)
Age 30-39	-0.009	0.024	-0.006	-0.000
	(0.014)	(0.013)*	(0.014)	(0.015)
Age 50-64	-0.008	0.005	-0.042	-0.048
	(0.017)	(0.014)	(0.014)***	(0.013)***
Maori	-0.029	-0.012	-0.025	0.018
	(0.017)*	(0.014)	(0.017)	(0.019)
Pacific	-0.070	-0.087	-0.018	-0.042
	(0.024)***	(0.021)***	(0.024)	(0.025)
Other	-0.059	-0.001	-0.000	-0.044
	(0.027)**	(0.025)	(0.032)	(0.042)
Ethnicity Not Specified	0.008	-0.069	0.086	0.039
	(0.059)	(0.061)	(0.107)	(0.090)
Married	0.070	0.052	0.044	0.004
	(0.027)***	(0.019)***	(0.020)**	(0.023)
No Qualifications	-0.059	-0.067	-0.062	-0.038
	(0.016)***	(0.015)***	(0.013)***	(0.017)**
School Qualifications	-0.009	-0.035	-0.008	-0.001
	(0.015)	(0.013)***	(0.013)	(0.013)
Degree	0.117	0.091	0.072	0.087
	(0.022)***	(0.018)***	(0.018)***	(0.026)***
Year	Yes	Yes	Yes	Yes
Household Type	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
Log Hourly Earnings	-0.426	-0.419	-0.422	-0.445
	(0.038)***	(0.045)***	(0.029)***	(0.032)***
Constant	1.243	1.218	1.208	1.339
	(0.117)***	(0.146)***	(0.085)***	(0.101)***
Observations	3577	4274	3964	4762
R squared	0.279	0.264	0.235	0.174

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: Individuals aged 20-64 in the first year, with positive hourly earnings in each year Estimated using Stata's `svyreg` command. Omitted categories are Age 40-49, European, Unmarried and Post-school non-degree qualifications. The dependent variable represents the natural logarithm of the ratio of second year to first year hourly earnings. The coefficients (βs) on categorical variables may be interpreted in terms of the percentage change in this ratio due to belonging to the particular category compared to belonging to the omitted category, according to the formula (exp(β)-1)*100.