Modelling wage dynamics among Australian workers

Tim Higgins
Australian National University

Key Words: Wages, HILDA, microsimulation, variance components, HECS

Accurate modelling of earnings dynamics is critical for projections of poverty dynamics and inequality, taxation revenue and savings, and public policy costings. Despite extensive analysis of earnings dynamics using the PSID, there has been little research using Australian longitudinal sources. This paper seeks to partially address this gap by using the first seven waves of HILDA to estimate and compare dynamic wage models. Residuals from a mean fit to hourly wage are decomposed into permanent and transitory components, and it is found that the variance of the components differ by employment state. A random walk for the permanent component can be justified on economic and empirical grounds, and the models are extended to allow for non-Gaussian errors and serial correlation in the transitory component. Short-term predictive ability of the models is mixed, and possible improvements are discussed. The implications of the models developed are illustrated through projections of outstanding debt and repayments through the Higher Education Contribution Scheme for a hypothetical population.

JEL classification: C23, C53, C54, I22

Acknowledgments
This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper are those of the author and should not be attributed to either FaHCSIA or the Melbourne Institute.

1. Introduction

There has been considerable econometric analysis of income dynamics through US and other international data sources (e.g., Lillard and Willis, 1978; MaCurdy, 1982; Meghir and Pistaferri, 2004; Geweke and Keane, 2000). Models of income dynamics can improve our understanding of poverty dynamics and inequality (e.g., Baker and Solon, 2003), and are critical inputs into microsimulation models for policy costing and development (e.g., Andreassen, et al., 1996; O'Donoghue, 2001). In contrast to international studies, there is very little published material on earnings dynamics in Australia.[1] This paper seeks to improve earnings projections for policy analysis in Australia through investigation and modelling of wage dynamics from HILDA. The layout of this paper is as follows. First, a cursory empirical analysis of HILDA earnings data mobility is presented. Modelling dynamic earnings requires multiple components: a model of labour force transitions, and a model of earnings conditional on labour force state. The focus of the current paper is on models of earnings, and hourly wage specifically. Empirical and theoretical aspects of earnings modelling in

[1] Two exceptions are Breusch and Mitchell (2003) and Keegan and Thurecht (2008).

the literature are summarised, with emphasis on permanent and transitory earnings components and stochastic model structures. This is used in the development of models of increasing complexity for hourly wage. The models are developed using the first seven waves of HILDA, and the predictive ability of the models are tested against observed hourly wage for Waves 8 and 9. For illustration of the implications of earnings model complexity, outstanding debt and compulsory repayments under the Higher Education Contribution Scheme debt are simulated for a hypothetical population.

2. Earnings mobility in HILDA

Exploratory data analysis of Australian earnings data is first undertaken in order to examine and identify the different levels of variability in earnings, and specifically individual variability over time. Earnings mobility can be defined '…as a change in individual ranks within a distribution' (Moffitt and Gottschalk, 1998: 18). Examination of earnings cross-sections over time may present a picture of relative stability, evidenced by low variability in distributional statistics, such as mean, median and other percentiles, and measures of inequality such as Gini coefficients. However, when one considers individual earnings over time, the stability disappears.

Earnings mobility in HILDA[2] is explored by decomposing male and female full-time and part-time earnings from wages and salary into quintiles, and measuring the proportion of individuals within each quintile in Wave 1 that remain in the same quintile, or move to other quintiles in subsequent waves. The full-time quintiles in each wave are derived from full-time wage and salary earners aged 20 to 60 in Wave 1 who have remained full-time in subsequent waves. Part-time quintiles have been similarly derived. Table 1 (A) and (B) report rates of earnings mobility for persons who have remained employed full-time. Table 1 (C) gives the earnings mobility rates for females who have remained employed part-time.[3] More complete results are presented in While mobility in earnings can arise due to labour force variation brought about by life course events, the discussion above is confined to earnings mobility within static labour force states. Earnings mobility of this nature is associated with changes due to promotion or demotion, industry or occupation.

Figure 1 and 2.

It is clear that there is considerable mobility for those who remain employed either full-time or part-time. For both males and females in the first (lowest) full-time quintile, there is almost a 40 per cent chance of moving to a higher quintile within one year. As expected, most who move only do so to the next highest group, however, close to 15 per cent move two or more quintiles. Despite the large proportion transitioning in the first year, the majority who started in the first quintile remain in that group six years later (in Wave 7), indicating strong persistency in earnings. The pattern for the fifth, or highest, quintile is similar, with a majority remaining in the

[2] The estimate of wages and salaries that is used here includes incorporated business income, but excludes unincorporated business income (see Watson, 2010: 51). For a description of the HILDA data see Watson, 2010.

[3] Mobility rates for males with part-time employment over the seven waves have not been included as they are particularly variable due to the low number of respondents in the survey with these characteristics.

highest quintile within one year, and a smaller though still large proportion persisting after 6 years. For those not at the extremes of the earnings distribution, there is scope for moving up or down and consequently persistency of earnings is not as marked as for the tails of the distribution.

TABLE.1 MOBILITY RATES ACROSS EARNINGS QUINTILES FOR FULL-TIME AND PART-TIME EMPLOYMENT. HILDA, WAVES 1 TO 7

(A) MALES, FULL-TIME EMPLOYED

	Wave 1 to Wave 2					Wave 1 to Wave 7				
Quintile	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
1st	0.63	0.24	0.10	0.02	0.02	0.54	0.25	0.09	0.09	0.03
2nd	0.24	0.51	0.17	0.07	0.02	0.29	0.38	0.19	0.10	0.04
3rd	0.08	0.19	0.49	0.21	0.03	0.10	0.25	0.38	0.17	0.11
4th	0.02	0.05	0.20	0.56	0.17	0.05	0.08	0.25	0.39	0.23
5th	0.02	0.02	0.05	0.14	0.77	0.01	0.04	0.10	0.24	0.60

(B) FEMALES, FULL-TIME EMPLOYED

	Wave 1 to Wave 2					Wave 1 to Wave 7				
Quintile	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
1st	0.60	0.26	0.07	0.04	0.04	0.48	0.26	0.19	0.04	0.03
2nd	0.19	0.51	0.24	0.06	0.01	0.21	0.46	0.20	0.11	0.02
3rd	0.08	0.15	0.51	0.24	0.02	0.18	0.19	0.33	0.21	0.09
4th	0.08	0.06	0.15	0.51	0.20	0.06	0.06	0.16	0.45	0.27
5th	0.05	0.03	0.03	0.16	0.73	0.08	0.03	0.11	0.19	0.59

(C) FEMALES, PART-TIME EMPLOYED

	Wave 1 to Wave 2					Wave 1 to Wave 7				
Quintile	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
1st	0.63	0.25	0.07	0.04	0.02	0.46	0.24	0.22	0.07	0.02
2nd	0.14	0.49	0.26	0.08	0.03	0.14	0.38	0.14	0.27	0.06
3rd	0.11	0.20	0.43	0.22	0.03	0.16	0.18	0.29	0.20	0.16
4th	0.06	0.04	0.19	0.50	0.21	0.11	0.14	0.26	0.27	0.22
5th	0.06	0.02	0.05	0.16	0.71	0.13	0.06	0.09	0.18	0.55

Table 2 gives the number of respondents in each category that remained either full-time or part-time employed over the seven survey waves.

TABLE.2. NUMBER OF SURVEYED AGES 20 TO 60 WHO HAVE MAINTAINED THE SAME EMPLOYMENT STATE.

	Wave 1	Waves 1 to 2	Waves 1 to 3	Waves 1 to 4	Waves 1 to 5	Waves 1 to 6	Waves 1 to 7
Males: FT employed	3,000	2,506	2,121	1,864	1,687	1,540	1,403
Males: PT employed	348	170	98	58	45	42	36
Females: FT employed	1,705	1,234	976	789	676	581	515
Females: PT employed	1,402	935	660	503	420	342	275

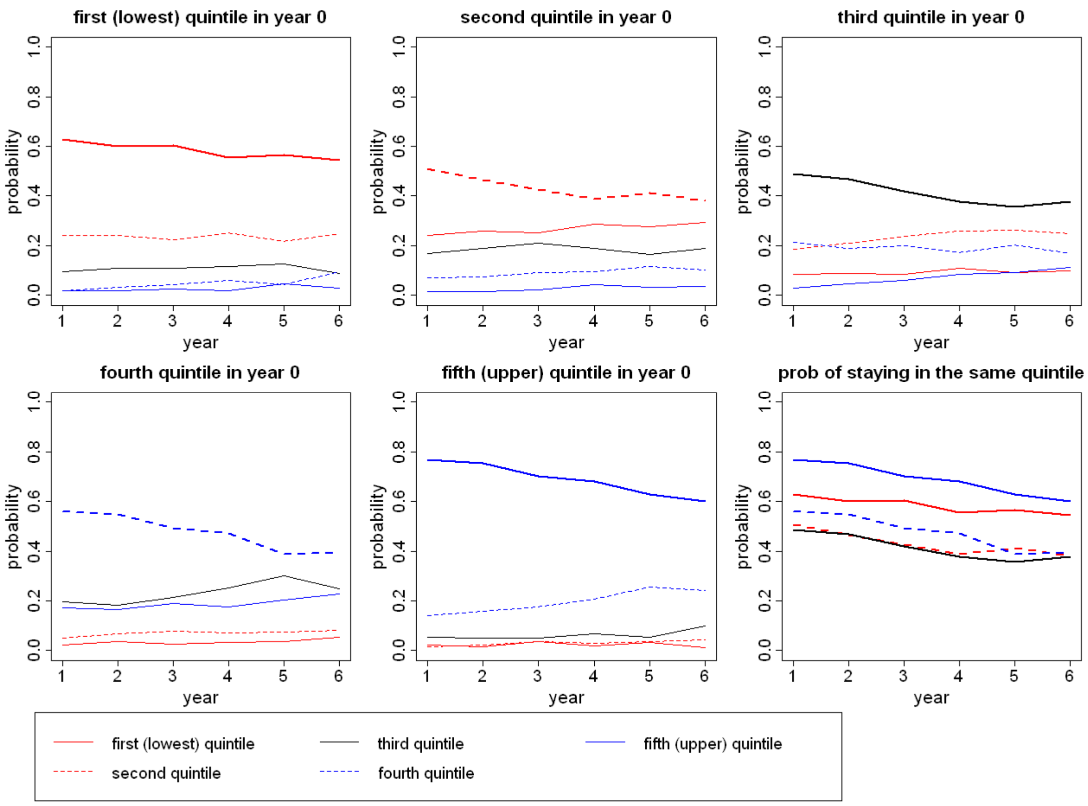
Notes: those with zero, negative, or missing earnings have been excluded.

The proportion moving two or more quintiles within one year ranges from 4 to 15 per cent depending on the starting group. This increases to over 25 per cent for some full-time groups after six years, with even greater mobility for part-time earners.

The large proportion moving one quintile, and the smaller yet still substantial proportion shifting two or more quintiles, is consistent with the international and domestic literature on income mobility. Moffitt and Gottschalk (1998) report numbers similar to those of Table 1 for US data.[4] They note that mobility at the upper and lower quintiles is smaller than in the middle quintiles. In an examination of British income mobility from 1991 to 1996, Jenkins (2000) shows that fewer than 60% of individuals remain in the same income group from year to year (where six groups are chosen based on weekly income). Breusch and Mitchell (2003) find a large degree of family income mobility in The Negotiating the Life Course Project data. Adjusting for household size and composition, they note that close to half of the sample remained in the same quintile over three years, 28 per cent moved to a higher quintile, 18 per cent moved to lower quintiles, while there was relative stability in the highest quintile. Keegan and Thurecht (2008) summarise the evidence for income mobility from Australia, observing that males generally experience greater mobility than females.

While mobility in earnings can arise due to labour force variation brought about by life course events[5], the discussion above is confined to earnings mobility within static labour force states. Earnings mobility of this nature is associated with changes due to promotion or demotion, industry or occupation.

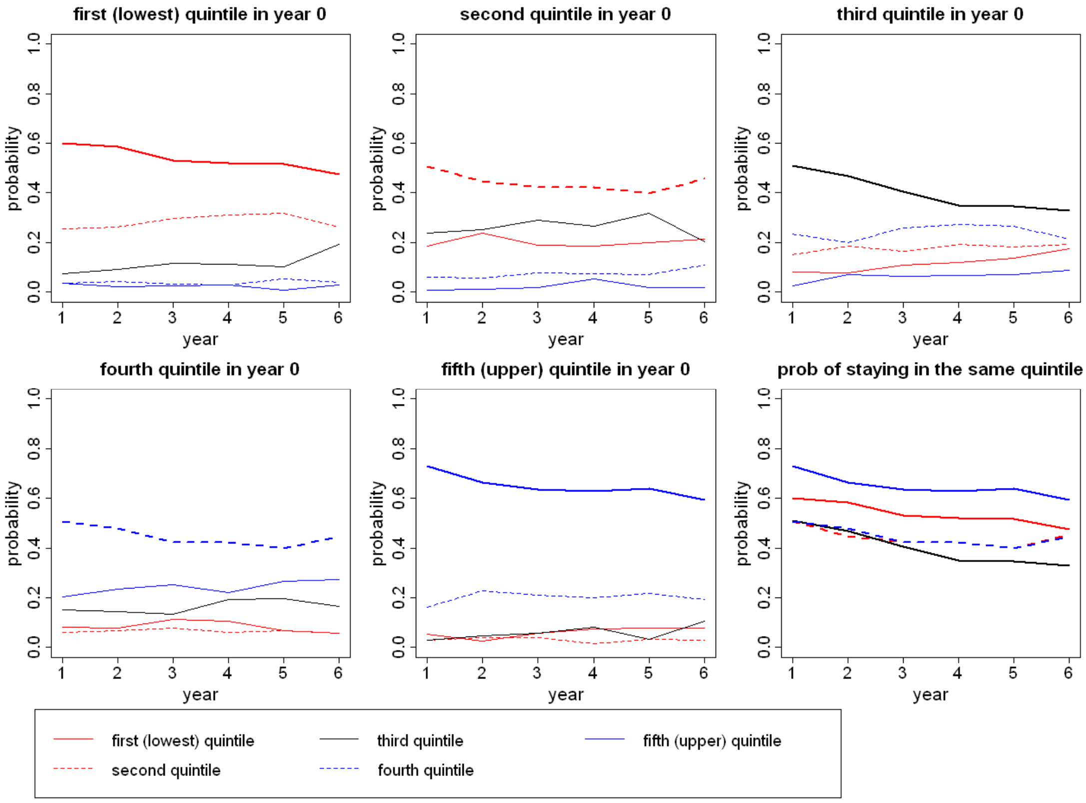
FIGURE 1 MALES, FULL-TIME EMPLOYED IN WAVE 1 (YEAR 0): MOBILITY ACROSS FULL-TIME EARNINGS QUINTILES



[4] They report percentages of 67, 21, 8, 3, and 1 for the one-year quintile mobility rates from the bottom quintile, and 7, 21, 44, 22 and 6 for the middle (third) quintile mobility rates.

[5] e.g.,The birth of a child leading to part-time employment or leaving the labour force; reduced child dependency leading to mother entering full-time employment; education leading to periods of interrupted employment; reductions in work hours as preparation for retirement, etc.

FIGURE 2 FEMALES, FULL-TIME EMPLOYED IN WAVE 1 (YEAR 0): MOBILITY ACROSS FULL-TIME EARNINGS QUINTILES



3. Modelling earnings – empirical evidence and theoretical considerations

Empirical studies have shown that observed characteristics (such as occupation, industry, and length of paid employment) explain a relatively small proportion of variability in earnings (e.g., see Swan, 1997). Unobserved differences can arise due to temporary variation, through illness, higher duties, bonuses, and overtime, or due to permanent differences, such as intellectual ability, drive and determination. Additionally, permanent unobserved shocks to earnings may arise due to job mobility and promotions or demotions (e.g., see Meghir and Pistaferri, 2004), and other incidents not accommodated by observed transitions in labour force or life states. Together, the temporary and permanent differences manifest as unexplained differences in earnings between individuals. Temporary differences and permanent shocks may also account for unexplained changes in earnings over time for the same individuals. The identification of multiple levels of earnings variability, some temporary and others permanent, is critical when choosing an appropriate model structure. The complexity in modelling earnings is increased further by noting that changes over time are also affected by productivity growth (i.e., real wage inflation) which will shape the rate of earnings, and business cycles, which may influence the number of hours worked.

The multiple layers of variability lend themselves to modelling through variance component models. Variance component models applied to dynamic earnings have been used by econometricians over the past 30 years, and have grown in complexity as panel data has expanded in duration. The complexity of the error structure dictates the complexity of the model. Broadly, there are two types of models for earnings dynamics that have been favoured by researchers in the field. The first, which can be referred to as 'profile heterogeneity', assumes that individuals follow person-specific earnings profiles such that their initial earnings and growth rates vary in a systematic way across individuals. This assumption is consistent with economic theory; for example, heterogeneity exists in human capital investments across similar individuals (Becker, 1962). The second model assumes that serial dependency is a consequence of a random walk for earnings (i.e., earnings has a unit root) rather than persistent profile heterogeneity. Empirical support exists for both processes, however Baker (1997) tests the two processes against long duration longitudinal earnings data and concludes that there is greater evidence for profile heterogeneity. More recently Alvarez *et al.* (2002) have argued that both processes should be accommodated in earnings models.

Lillard and Willis (1978) were among the first to recognise the need to model earnings dynamically. They applied a standard earnings function consisting of observed variables such as education, length of work experience, and gender to the Panel Study for Income Dynamics (PSID). [6] Unobserved variables were represented by a random component that captured heterogeneity in permanent differences between individuals, and a stochastic (transitory) error component that incorporated serial correlation through an AR(1) process. This was followed by MaCurdy (1982) who concluded that an ARMA(1,2) or ARMA(2,1) process is more appropriate for the transitory error component. He allowed for persistency in earnings through a unit root, modelling the first difference of the natural log of earnings against covariates. Variance component structures have since increased in complexity. Moffitt and Gottschalk (2002) decomposed PSID male annual earnings from 1969 to 1996 into permanent and transitory components, and fit a dynamic earnings model where the permanent component followed a random walk and varied with age, and the transitory error was modelled as an ARMA(1,1) process. They allow for an initial permanent differential at the start of the working life (due to unobservable factors such as intellectual ability, drive and ambition, and skill set) which captures variation between like individuals and is akin to profile heterogeneity. By imposing a unit-root process on the permanent unobserved component the initial profile heterogeneity is maintained, while allowing for future variation due to permanent shocks. Their approach was broadly followed by both Dickens (2000) and Ramos (2003) for Great Britain.

Recent empirical work has suggested that random shocks in dynamic earnings models are non-normal, and modern statistical methods have allowed application of more sophisticated stochastic processes (e.g., see Geweke and Keane (2000); Hirano (2002)). Furthermore, heteroscedasticity in the variance of earnings shocks has been allowed for by some researchers (e.g., Chamberlain and Hirano (1999), Meghir and Pistaferri (2004)).

---

[6] The PSID is the most widely used dataset among researchers into income dynamics in the United States. This is a longitudinal survey of the civilian non-institutional US population since 1968.

In summary, while there has been general acceptance in the literature that the error term should be separated into transitory and permanent components, there is debate as to the appropriate process for each component, including the choice of conditional mean and variance. The choices taken should ideally be consistent with economic theory, and it is critical that they should be informed by the observed data. Importantly, when Alvarez *et al.* (2002) fit earnings models allowing for heterogeneity to both Danish data and the PSID, completely different processes were selected for the two data sets, suggesting that 'one size fits all' does not apply when it comes to dynamic earnings models.[7]

4. Developing models of hourly wage

When modelling full-time annual earnings it is arguable whether predictability will be improved by decomposing earnings into a wage rate and number of hours worked. In contrast to full-time earnings, there is considerable volatility in part-time earnings as a consequence of variability in the number of hours worked per week. Further, the socio-economic determinants of wage rate and hours worked differ. For modelling purposes it may be convenient to express annual earnings in terms of the number of hours worked, or number of weeks worked per annum. Under this formulation, annual earnings become:

$$\text{Annual Earnings} = \text{hourly earnings} \times \text{hours per week} \times \text{weeks per year}$$

While hourly wage is a superior measure of earnings for individuals who work irregular hours week to week but are paid a fixed rate, weekly earnings is superior for individuals who may work varying hours but have a fixed annual salary (Keegan and Thurecht, 2008). The choice of whether to derive equations for weekly earnings or hourly earnings depends on how labour force state is produced in the modelling. Part-time status typically manifests as a reduction in the number of hours per week worked, rather than a reduction in the number of weeks of full-time work per annum. Under this assumption hourly wage is a more appropriate earnings measure than weekly earnings.[8]

In this section models of hourly wage are developed. Hourly wage is extracted from HILDA by dividing weekly gross wages and salaries (in all jobs) by the number of hours worked per week (in all jobs) (Watson, 2010). Hourly wage is first adjusted by inflating to 2007/08 values with Average Weekly Ordinary Time Earnings (AWOTE) (ABS, 2010).

A review of earnings patterns in Australia is given by Keegan and Thurect (2008). Determinants reviewed include gender, age, education, marital status and family composition, immigrant status, industry and occupation, and health status. The

[7] A limitation of variance component modelling in this context is that studies to date have been confined to a homogenous group, namely employed male workers. As noted by Jenkins (2000), the variance component approach is not well suited to taking into account permanent changes in labour force or life stage (such as unemployment or shifts to part-time work). A way of mitigating this restriction is to fit separate models of earnings to each labour force state. Earnings functions can then be developed which account for the particular labour force circumstances facing each individual.
[8] This is consistent with the approach taken for the NATSEM microsimulation model APPSIM (Keegan and Thurecht, 2008).

authors also note the possible importance of experience, although they acknowledge the difficulty in quantifying this measure. There is also evidence of interactions; for example, those with a university degree have a greater increase in wage rates with age than those without a degree (Kalb and Scutella, 2002). The following covariates were considered in models developed in this paper: age last birthday, sex, education, marital status, age of youngest child, part-time or full-time employment status, and experience (being the number of years of employment).[9] While the use of these covariates is supported on theoretical grounds (e.g., see Keegan and Thurecht, 2008), they are also key variables used among developers of microsimulation models.

Part-time/full-time employment status as a potential covariate deserves some discussion prior to modelling. In the HILDA data it is found that the hourly wage for full-time employed is greater than part-time employed for both sexes across the deciles of the wage distribution. This is consistent with the notion of the full-time wage premium whereby higher wages are paid to those working full time.[10] Analysis of the first seven waves of HILDA reveals that male full-time wages are 29.5 per cent more than male part-time wages, while female full-time wages are only 7.8 per cent more than female part-time wages. Both mean and decile male part-time wages are less than the corresponding values for part-time females. However, when we exclude those with youngest ages from the analysis (below age 22), the spread reduces to 11.3 per cent for males and 1.5 per cent for females. This difference can be attributed to lower hourly wage rates among young persons, the majority of whom are employed part-time; Rodgers (2004) finds that after controlling for job type and worker-specific characteristics (such as age, education and occupation), there is no statistically significant difference in hourly wage between part-time and full-time workers in Wave 1 of HILDA. While age and education (and interactions) are controlled for in the wage models that follow, occupation is not; hence, there is the possibility that employment state remains correlated with hourly wage. Consequently, full-time and part-time workers are considered separately when developing models for hourly wage in this chapter.

Although hourly wage is clearly a function of lagged hourly wage, this dependency was not modelled by including lagged log hourly wage as a regressor. Alternatives include first differencing (under the assumption that earnings follow a unit root process), or by capturing the heterogeneity through inclusion of a permanent error component that varies across individuals (profile heterogeneity). As discussed above, the assumption of a unit-root for earnings or the assumption of profile heterogeneity need not be exclusive of one another; by imposing a unit-root process on an initial permanent unobserved component, profile heterogeneity can be accommodated while allowing for future variation through temporary and permanent shocks. This is the approach taken by Harris and Sabelhaus (2003) and is used for models M.3 to M.5

[9] Strong correlation between age and experience (>0.9) led to the omission of experience as a covariate from the models considered. Including experience as a covariate in addition to age (and alternately in place of age) had negligible marginal impact on the explanatory power of the mean fit.

[10] See Lettau (1994) for a discussion of the theory behind the full-time wage premium. Explanations for the premium include the possibility that part-time workers are less productive and therefore command lower salaries, or that poorer paying jobs tend to be disproportionately part-time. Lettau (1994) estimates a 16 percent premium for full-time workers, and Harris and Sabelhaus (2003) use 15 percent in their model of hourly wage.

below.[11] On the back of the theoretical and empirical evidence above, permanent and transitory components should be incorporated into the earnings models. Individuals have initial earnings that can be decomposed into a mean component explained by covariates such as age, gender, and education, and an unobserved differential. The differential between actual and mean earnings is assumed to be the result of unobserved covariates[12], temporary shocks, and the cumulative effect of permanent shocks. Of critical importance are the dynamic processes for earnings, namely, the stochastic processes that determine the shocks to hourly wage.

As a first step, those observations with missing dependent and independent variables were excluded from the analysis. Next, those individuals receiving less than \$5 per hour or more than \$120 per hour, constituting approximately 2 per cent of the observations, were excluded. Hourly wage rates for those aged up to 21 are considerably lower than subsequent ages; these were omitted for the purpose of the analysis and modelling, as were those aged 65 and above. This reduced the sample to 8,946 individuals and 37,884 observations.

A model of hourly wage rates is first fit that ignores the known dependence structure in wages. The logarithm of hourly wage is regressed on age, education, part-time or full-time employment status, marital status and presence and age of youngest dependent children. Separate models were fit for male and females. A natural spline component for age is included with 3 degrees of freedom and an interaction was included between age and education. The model can be written:

$$\log\left(W_{it}\right) = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}\,, \qquad \text{(M.0)}$$

where W_{it} is the hourly wage for individual i in year t. The coefficient estimates and their significance levels are given in Table 3. The first stage regression provides an estimate of $\boldsymbol{\beta}'$ which remains unbiased in the event of heteroscedasticity or serial correlation in the residuals. The explanatory power of the covariates is low, producing R-squared values of 0.14 and 0.16 for males and females respectively.[13] While full-time/part-time status is only significant for females at the 10% level, it is found subsequently that the between and within person variation in hourly wage differs markedly between full-time and part-time workers. Therefore, employment status is considered as a factor when modelling the variance and covariance of earnings. Decomposition of the residuals is performed to ascertain the existence and structure of transitory and permanent variance components. This approach is consistent with those of Moffitt and Gottschalk (1998) and Ramos (2003), among others.

[11] In contrast to Harris and Sabelhaus (2003) who modelled earnings for all workers and considered iid normal shocks, here hourly wage for part-time and full-time workers are considered separately, and models are developed which allow for non-Gaussian errors among other processes.
[12] Occupation and industry are omitted from the mean effect, and hence are treated as unobserved. Although these are determinants of wage, for the majority of an individual's career they typically remain fixed, and therefore can be subsumed into the unobserved permanent component in the earnings model.
[13] Adding additional interactions between covariates has only a minor impact on the fitted models, and has a negligible effect on the residual variances and autocovariances.

TABLE 3 COEFFICIENTS FOR MODEL M.0. LOG HOURLY WAGE REGRESSIONS FOR MALES AND FEMALES.

Independent variables	Males		Females	
	Estimate	Standard error	Estimate	Standard error
Intercept	3.089***	0.023	3.070***	0.017
Age (df =1)	0.261***	0.027	0.185***	0.022
Age (df =2)	0.841***	0.057	0.629***	0.044
Age (df =3)	0.128***	0.031	0.121***	0.030
Marital status (not married)	-0.079***	0.008	-0.052***	0.006
Dependent children (youngest 5 or under)	0.000	0.010	-0.015	0.010
Dependent children (youngest 6 or over)	-0.005	0.009	-0.057***	0.007
Education (tafe)	-0.084***	0.029	-0.167***	0.024
Education (<=yr12)	-0.099***	0.028	-0.169***	0.022
Employment status (part-time)	-0.068***	0.011	0.019*	0.006
Age (df = 1) x education (tafe)	-0.097***	0.033	-0.083**	0.032
Age (df = 2) x education (tafe)	-0.446***	0.072	-0.307***	0.064
Age (df = 3) x education (tafe)	-0.039	0.038	-0.097**	0.041
Age (df = 1) x education (<=yr12)	-0.204***	0.035	-0.082***	0.028
Age (df = 2) x education (<=yr12)	-0.589***	0.069	-0.430***	0.058
Age (df = 3) x education (<=yr12)	-0.081	0.039	-0.133***	0.036

*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

A naive stochastic model is one that assumes that all the earnings variability is due to transitory variation, and that permanent differences don't exist:

$$r_{it} = \upsilon_{it}, \qquad \text{(M.1)}$$

where υ_{it} is the serially uncorrelated transitory component with mean 0 and variance σ_{υ}^{2}. In Table 4 the variances and standard errors are given in the diagonal, autocovariances and standard errors are given below the diagonal, and autocorrelations are presented above the diagonal.[14]

It is clear from Table 4 that M.1 should be rejected - there are obvious deficiencies due to the presence of serial correlation. Further, it appears that autocorellations for males are much higher than for females, likely due to the higher rates of persistent full-time employment for males and fewer incidents of transitions between part-time and full-time employment states. This was explored further by decomposing the residual variances and autocovariances by full-time and part-time employment status. Examination of these results with respect to period suggests the values are generally stationary over the first seven waves of HILDA, and so were aggregated over all years in Figures 3 and 4.[15] It is apparent that the variances for part-time males and females

[14] Standard errors were generated by using the block bootstrap procedure applied to the observed residuals Bootstrapping was performed via the R function 'tsboot' in the library 'boot' with a fixed block length of seven corresponding to the number of waves of data, and 500 bootstrap resamples. This ensures that heteroscedasticity and serial correlation are accounted for.

[15] To test whether the residual variances and autocovariances are distorted by period effects, year was included as a factor in model M.0. Including year led to a negligible difference in the variances and

are considerably greater than for full-time males and females.[16] Consequently, all further models treat full-time and part-time employment states separately.

TABLE 4 MALES AND FEMALES, VARIANCES, AUTOCOVARIANCES AND AUTOCORRELATIONS FOR RESIDUALS FROM MODEL M.0, AGES 25-64.

MALES

Year	2001	2002	2003	2004	2005	2006	2007
2001	0.197 (0.009)	0.68	0.63	0.62	0.58	0.58	0.55
2002	0.118 (0.012)	0.187 (0.009)	0.74	0.67	0.63	0.61	0.60
2003	0.106 (0.009)	0.121 (0.012)	0.181 (0.01)	0.72	0.66	0.64	0.60
2004	0.112 (0.004)	0.115 (0.008)	0.122 (0.012)	0.199 (0.01)	0.72	0.69	0.64
2005	0.102 (0.004)	0.108 (0.004)	0.113 (0.008)	0.129 (0.012)	0.201 (0.010)	0.73	0.68
2006	0.106 (0.008)	0.108 (0.004)	0.111 (0.004)	0.128 (0.008)	0.132 (0.012)	0.214 (0.01)	0.74
2007	0.096 (0.013)	0.102 (0.008)	0.102 (0.004)	0.113 (0.004)	0.120 (0.008)	0.137 (0.012)	0.195 (0.01)

FEMALES

Year	2001	2002	2003	2004	2005	2006	2007
2001	0.150 (0.004)	0.53	0.53	0.47	0.47	0.41	0.40
2002	0.070 (0.009)	0.157 (0.004)	0.55	0.52	0.51	0.43	0.37
2003	0.068 (0.008)	0.077 (0.009)	0.148 (0.004)	0.63	0.50	0.53	0.48
2004	0.059 (0.004)	0.068 (0.008)	0.079 (0.009)	0.149 (0.004)	0.56	0.57	0.50
2005	0.060 (0.004)	0.066 (0.004)	0.065 (0.009)	0.072 (0.009)	0.145 (0.004)	0.59	0.54
2006	0.054 (0.008)	0.056 (0.004)	0.069 (0.004)	0.073 (0.008)	0.079 (0.009)	0.153 (0.004)	0.62
2007	0.053 (0.009)	0.047 (0.008)	0.061 (0.004)	0.062 (0.004)	0.067 (0.008)	0.083 (0.009)	0.144 (0.004)

autocovariances, and there was an absence of any trend in mean log hourly earnings. Hence, year was excluded from the regression models.

[16] Results were also extracted for individuals who transitioned between part-time and full-time states (and vice versa) and it was found that autocorrelations were very low.

FIGURE 3 RESIDUAL VARIANCES AND AUTOCOVARIANCES FOR ALL YEARS, 2001 TO 2007. FULL-TIME AND PART-TIME WORKERS CONSIDERED SEPARATELY.

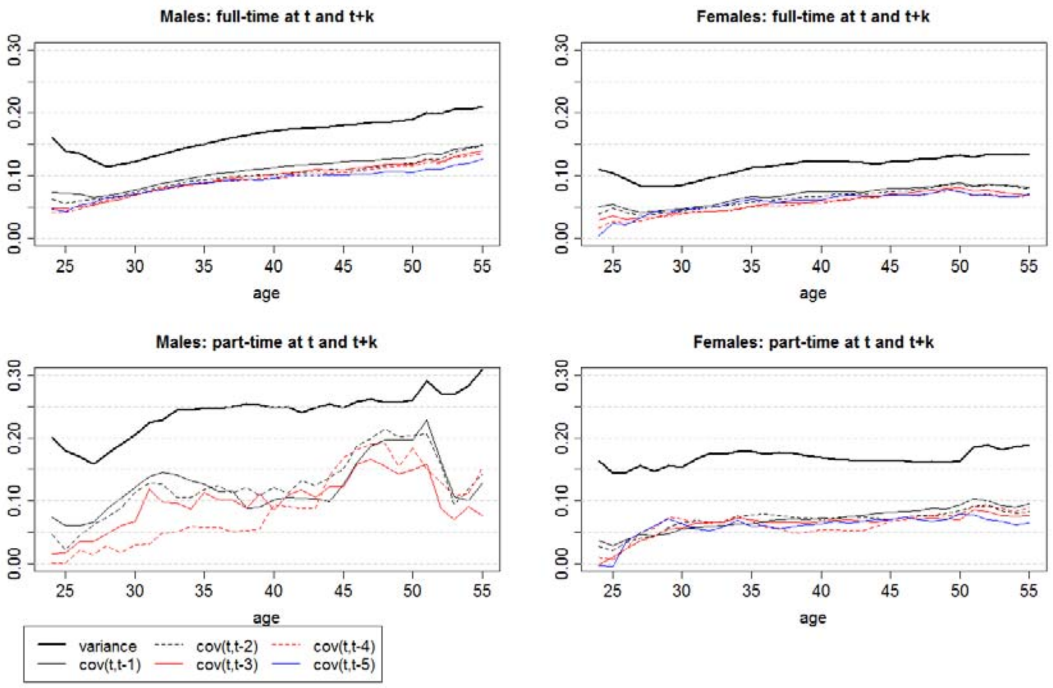
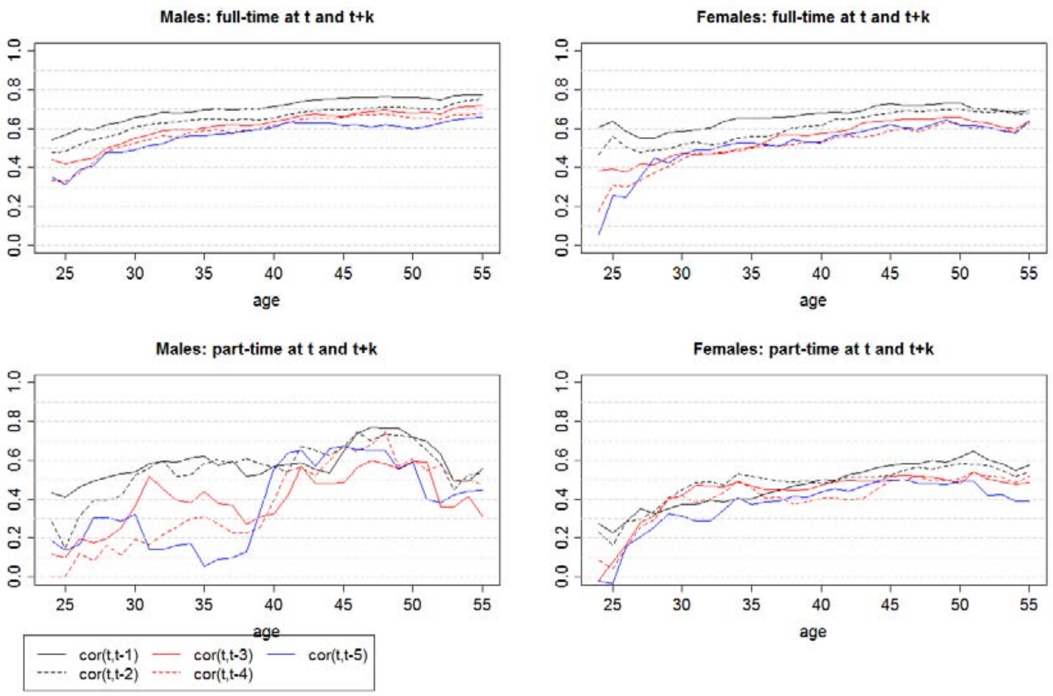


FIGURE 4 RESIDUAL AUTOCORRELATIONS FOR ALL YEARS, 2001 TO 2007. FULL-TIME AND PART-TIME WORKERS CONSIDERED SEPARATELY.



A naive deterministic model is one that assumes that all earnings variability is due to permanent differences and there is no allowance for earnings mobility, however, in this circumstance one would expect autocorrelations to be close to unity. While one could attempt to allow for serial dependency by fitting a time series process for the transitory error component, economic theory and the earnings literature is strongly supportive of decomposition into a permanent and transitory component. The simplest model is the canonical permanent-transitory model, consisting of a white noise transitory and time-invariant permanent component:

$$r_{it} = u_i + \upsilon_{it}, \qquad \text{(M.2)}$$

where u_i is the time-invariant component, and υ_{it} is as for model M.1. Values for the components of variance for M.2 can be found by extracting the unobserved earnings shocks (the differenced residuals, Δr_{it+k}). The variance of the transitory component is then easily estimated (see Appendix A1) with the remaining variance being attributed to the permanent variation between individuals, σ_u^2. The estimated values for males and females for ages 25 to 54 and over all years are given in Table 5.

TABLE 5 VARIANCE COMPONENT ESTIMATES FOR MODEL M.2. AGES 25 TO 54

	Variance estimate	All	FT at *t* and *t*+*k*	PT at *t* and *t*+*k*	FT at *t* and PT at *t*+*k*	PT at *t* and FT at *t*+*k*
Males	Total variance	0.175	0.168	0.25	0.181	0.266
	σ_υ^2	0.069	0.063	0.110	0.097	0.119
	σ_u^2	0.106	0.105	0.140	0.084	0.147
Females	Total variance	0.141	0.118	0.171	0.126	0.173
	σ_υ^2	0.071	0.048	0.087	0.092	0.088
	σ_u^2	0.070	0.070	0.084	0.034	0.085

M.2 allows for serial correlation through the permanent component, however, it is inconsistent with the observations of Figures 3 and 4; the observed residual variances and covariances for full-time males and females clearly increase with age. The pattern of increasing variance, and long lagged autocovariances with age, can be captured by modelling the permanent component as a random walk:

$$r_{it} = u_{it} + \upsilon_{it}, \qquad u_{i,t+1} = u_{it} + w_{it} \qquad \text{(M.3)}$$

where, w_{it} represents an i.i.d. random permanent shock with mean 0. As before, u_{it} represents unobserved differences in hourly wage due to factors such as IQ, drive and ambition, and occupation and industry, among others. Many of these may have a permanent effect on an individual's hourly wage, while others will be less permanent, though have long persistence (such as occupation and industry of employment[17]). Changes in earnings from year *t* to year *t*+1 occur due to changes in $\mathbf{x}_{it}$, but also due to random shocks to both the unobserved permanent and transitory components. The

[17] While changes in occupation and industry may clearly affect hourly wage, employment type will also have a bearing on the rate of growth of wages. Individuals will have different rates of earnings growth which have a persistent effect on wage. Incorporating a random shock to the permanent component allows for this type of variation.

random shocks affect the transitory income through υ_{it} and the permanent component of earnings through the permanent shock w_{it}.[18] The steps used to estimate the variance of the shocks, w_{it} and υ_{it}, are given in Appendix A1[19] and results are in Table 6.

TABLE 6 ESTIMATES OF THE VARIANCE OF PERMANENT AND TRANSITORY SHOCKS FOR MODEL M.3, AGES 25 TO 54.

	Variance estimate	FT at *t* and *t+k*	PT at *t* and *t+k*	FT at *t* and PT at *t+k*	PT at *t* and FT at *t+k*
Males	σ_w^2	0.006	0.000	0.007	0.003
	σ_υ^2	0.049	0.123	0.097	0.115
Females	σ_w^2	0.001	0.014	0.005	0.016
	σ_υ^2	0.043	0.056	0.083	0.052

When generating a population for earnings projections from M.3, one also needs an estimate of the initial permanent variance prior to shocks. Since,

$$\mathrm{var}(u_{i,t+k}) = \mathrm{var}(u_{it}) + \sum_{s=t}^{t+m-1} \mathrm{var}(w_{is}) = \mathrm{var}(u_{it}) + m\sigma_w^2,$$ [20]

the total variance for individual *i* at time *t* will be: $\mathrm{var}(r_{it}) = \sigma_{u_b}^2 + k\sigma_w^2 + \sigma_\upsilon^2$, where at the start of an individual's working life (at age *b*) there is an initial or base permanent differential in hourly wage, and where *k* is the difference between age at time *t* and age *b*. The variance of the base permanent component can be estimated by noting that the total residual variance at the start of an individual's working life (prior to permanent shocks) will equal $\mathrm{var}(r_{it}) = \sigma_{u_b}^2 + \sigma_\upsilon^2$. When predicting for time *t+1* for a population with existing earnings at time *t*, we need a starting estimate for the permanent unobserved earnings component, u_{it}. Unbiased estimates of the starting value for the permanent component are the residuals of the OLS fit, weighted to ensure that the variance of the permanent component is consistent with the variance as determined through the parameter estimates.

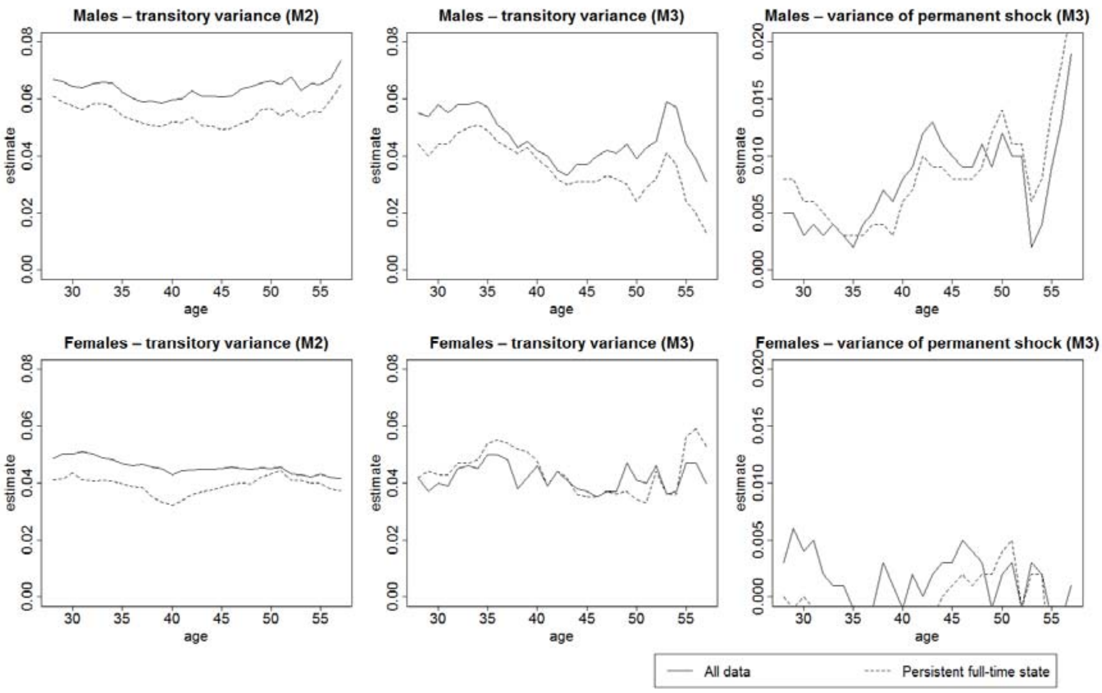
[18] While the term 'permanent' is used for consistency with the earnings literature, it is assumed to vary over time and can thus be considered a non-mean reverting effect as opposed to a strictly permanent effect (e.g., Dickens, 2000).

[19] An implication of assuming constant variance of earnings innovations with respect of log hourly wage is that higher levels of hourly wage will be subject to larger wage shocks than lower levels. To ascertain whether this is a realistic assumption, the variances of the one-step earnings innovations were extracted for full-time males, and both full-time and part-time females, for different log hourly wage deciles. The empirical variances across different deciles show no discernable pattern or trend, supporting the assumption of constant variances with respect to log hourly wage innovations.

[20] To be precise we could drop the notation for time *t* in the subscripts for the permanent component, and replace this with age *a*. Correlation and variance structures explored in HILDA appear to essentially display stationarity, yet there is evidence of variation with age. If we consider that a starting permanent differential exists at the commencement of an individual's labour force experience at age *b* due to persistent unobserved effects, then by age *a* (where $a > b$), the individual will have experienced $a - b$ potential permanent shocks. While we retain the notation *t* for the equations here, the random walk is intended to capture variation over the life course as an individual ages, rather than period effects.

It is worth pausing to discuss Tables 5 and 6. First, the values reported give a false impression of parameter stability with age. In Figure 5 the variance components are plotted by age for full-time individuals at *t* and *t*+*k*. In particular, this shows the parameter variability for males under M.3. Second, the parameter estimates have been derived from employment states at discrete times *t* and *t*+*k*, but have not required persistence in these states during intermediate periods. For comparison, variance component estimates were estimated for individuals who persisted at the full-time state for all intermediate waves between *t* and *t*+*k*. This is evidence that duration within existing employment state affects the variance components for hourly wage, a point of note for future earnings model development.

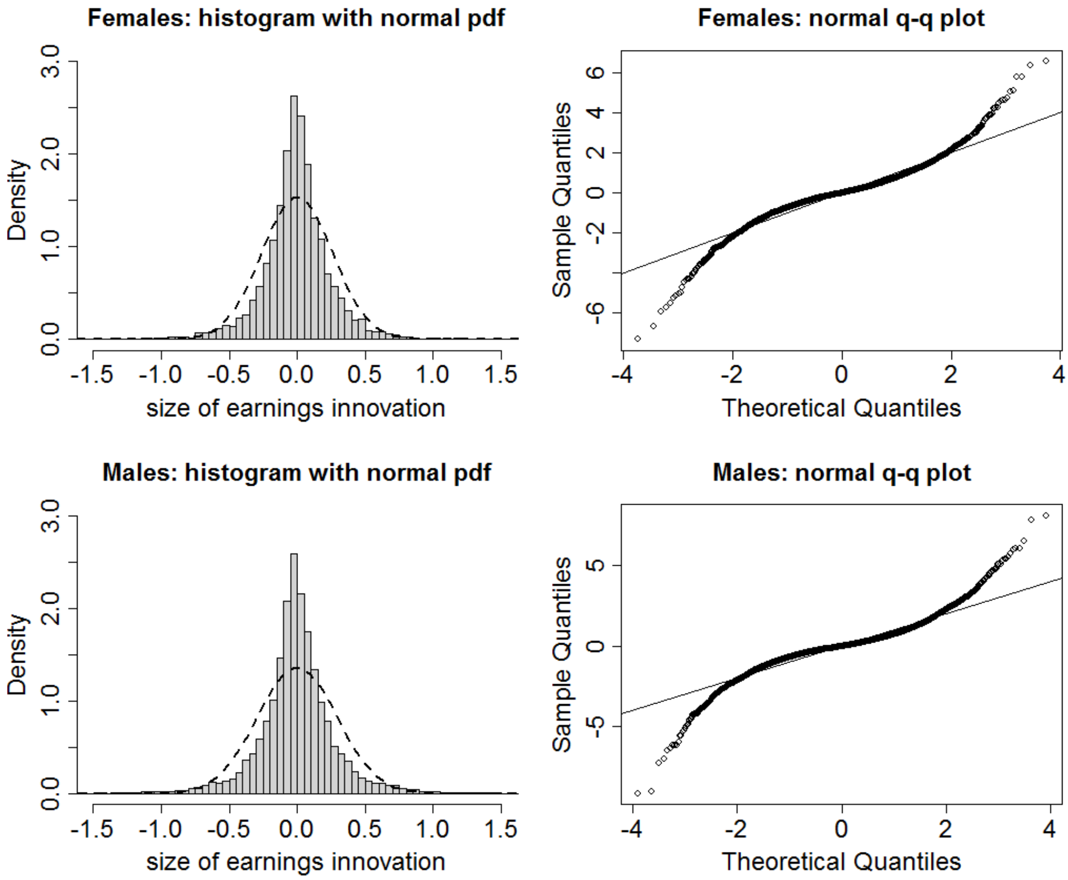
FIGURE 5 VARIANCE COMPONENT ESTIMATES BASED ON FULL-TIME AT *T* AND *T*+*K*. MODELS M2 AND M3. ALL AND PERSISTENT EMPLOYMENT STATES



Model M.3 can be extended by relaxing the assumption of normally distributed earnings shocks. Histograms and qq-plots for the unexplained earnings innovations clearly display non-normality[21]. While not plotted here, the earnings innovations for those remaining in the part-time state displays a similar level of kurtosis.

[21] See Appendix A1 for a definition of earnings innovation in this context. Normality was also formally rejected through a Shapiro-Wilk test.

FIGURE 6 ONE-STEP EARNINGS INNOVATIONS FOR AGES 25 TO 54. FULL-TIME AT *T*, FULL-TIME AT *T+1*.



To allow for non-normality, smoothed density curves were fit to the observed one-step earnings innovations. It was assumed that the distributions of the earnings innovations were consistent in shape for both the permanent and transitory shocks.[22] The resulting model, M.4, is identical to M.3 with the exception of the distribution of shocks, such that w_{it} and υ_{it} follow the shape of the smoothed empirical distributions in Figure 6.

Model M.3 can also be extended to allow for possible serial correlation in the transitory component:

$$r_{it} = u_{it} + \upsilon_{it}\,, \qquad u_{i,t+1} = u_{it} + w_{it}\,, \qquad \text{(M.5)}$$

$$\upsilon_{it} = \rho\upsilon_{i,t-1} + \varepsilon_{it}\,, \text{ or } \upsilon_{it} = \varepsilon_{it} + \theta\varepsilon_{i,t-1}\,.$$

[22] The density curves for transitory and permanent innovations were extracted for full-time to full-time, and part-time to part-time workers by weighting the observations for total innovations for each case by a constant factor such that the variances of the weighted densities were identical to the permanent and transitory variances extracted earlier. This approach is reasonable under the assumption that the transitory and permanent innovations are independent.

While the transitory component is written here as either an AR(1) or MA(1) process[23], AR(1) has been the preferred process for the transitory component in the earnings literature[24], with mixed empirical support from analysis of the PSID. As before, the variance of the earnings innovations can be used to solve for the parameters. In this case the covariance term for the transitory components is non-zero, and we have:

$$\operatorname{var}\left(r_{i,t+k}-r_{it}\right)=k\sigma_w^2+2\sigma_\upsilon^2-2\operatorname{cov}(\upsilon_{i,t+k},\upsilon_{it})=k\sigma_w^2+2\sigma_\varepsilon^2\left(\frac{1-\rho^k}{1-\rho^2}\right).$$

For individuals in full-time employment at t and $t+k$ the estimates for σ_ε^2 and ρ are statistically significant, being 0.040 and 0.25 respectively for females and 0.046 and 0.17 for males.[25] The low value for ρ indicates little persistence; the covariance between the current and lagged transitory error is only approximately 1/4 the transitory variance for females, and the covariance for lags beyond the first are essentially zero. It appears that by assuming a random walk for the permanent component, the serial dependency in earnings identified in Figures 3 and 4 is sufficiently captured.[26]

5. Model assessment

The prediction capability of the models was assessed by comparing hourly wage predictions for Waves 8 and 9 with the observed HILDA values. For each individual in each year of the simulation, $\boldsymbol{\beta}'\mathbf{x}_{it}$ was calculated from M.0. Unobserved components were then simulated using the processes and parameters estimated in the previous section, and added to the mean term. Projections were produced for individuals who were full-time in Wave 7 and Waves 8 and 9. The root mean-square prediction error (RSMPE) is used as a test of predictive accuracy. Results are given in Table 7 for full-time workers.

TABLE 7 ROOT MEAN-SQUARE PREDICTION ERRORS

	Males		Females	
Model	Wave 8	Wave 9	Wave 8	Wave 9
M.1	17.0	16.9	12.8	12.9
M.2	10.7	10.9	8.3	8.7
M.3	10.6	11.2	8.0	8.4
M.4	11.1	11.7	8.3	8.7

Values for M5 were the same as for M3 and so have not been included in Table 7; while there is evidence of autocorrelation in the transitory component for full-time workers, the effect has negligible impact on short-term earnings projections.

[23] While higher-level processes may be more suitable for the transitory component, the current analysis was based on an insufficient number of data periods to allow for processes that would require additional parameter estimates.
[24] e.g., see Lillard and Willis (1978), Chamberlain and Hirano (1999), Geweke and Keane (2000), Ramos (2003), Baker and Solon (2003).
[25] Estimates for ρ for part-time employment are not significant.
[26] An MA(1) process was also tested and it was found to be statistically significant for full-time males and females, however, as with the AR(1) process, the estimate was not materially significant.

There is a trade-off in model development, as additional model complexity does not necessarily reduce prediction error. RMSPE values indicate that the non-parametric error distribution for M.4 may overfit compared with M.3. Over the short prediction period considered, the accuracy of the predictions under Model M.2 is comparable with models M.3 and M.4 which incorporate permanent shocks. One would expect differences to become more apparent as the projection period increased. In the next section the implications to policy costing are illustrated through a simulation of HECS costs for a hypothetical population, and it is shown how the differences between M.2 and the more complex models can manifest over longer projection periods.

6. A policy example - implications to HECS costs

An area of government policy that can clearly benefit from models that incorporate earnings mobility is the Higher Education Contribution Scheme, or HECS. Under HECS, compulsory repayments are contingent on income. Specifically, the rate of repayment increases as income rises. If income falls below a minimum repayment threshold, no contributions are required, so HECS offers the attraction of default insurance. An implication is that without earnings mobility, individual debtors will be projected to either eventually repay all of their debt (if they lie above the minimum threshold) or none (if they lie below).

To illustrate the impact of the earnings models on HECS costs, HECS debt was calculated over a 30-year projection period for a subpopulation of females aged 25 to 35 as at the start of the HILDA survey. An initial debt of $20,000 and 2008–09 HECS repayment thresholds and rates are assumed (see Table 8).[27] HECS costs were simulated for a 30-year period assuming all debtors have static full-time employment (i.e., variation in earnings is not due to mobility between labour force states).[28] Hourly wages follow models M.1 to M.4 up to age 60 and were assumed to be below the lowest repayment threshold thereafter, and full-time individuals were assumed to work 40 hours per week (being the approximate average from the HILDA data over the first seven waves). Earnings in each future year were iteratively predicted via Monte Carlo simulation. Compulsory repayments and outstanding debt were calculated based on the simulated earnings.

TABLE 8 HECS REPAYMENT THRESHOLDS AND RATES FOR USE IN SIMULATIONS

Repayment threshold	Repayment rate (per cent)
Below $41,595	Nil
$41,595-$46,333	4.0
$46,334-$51,070	4.5

[27] Additionally, the projected incomes and income thresholds are increased with average growth in earnings at 4 per cent per annum, and outstanding debt is indexed at CPI, assumed to be 2.5 per cent per annum. 2.5 per cent is the middle of the Reserve Bank of Australia's acceptable band for price inflation, and 4 per cent is the approximate wage growth rate over the past decade in Australia (RBA, 2009a, 2009b). It was further assumed that education and marital status remained constant for each individual for the duration of the projection.

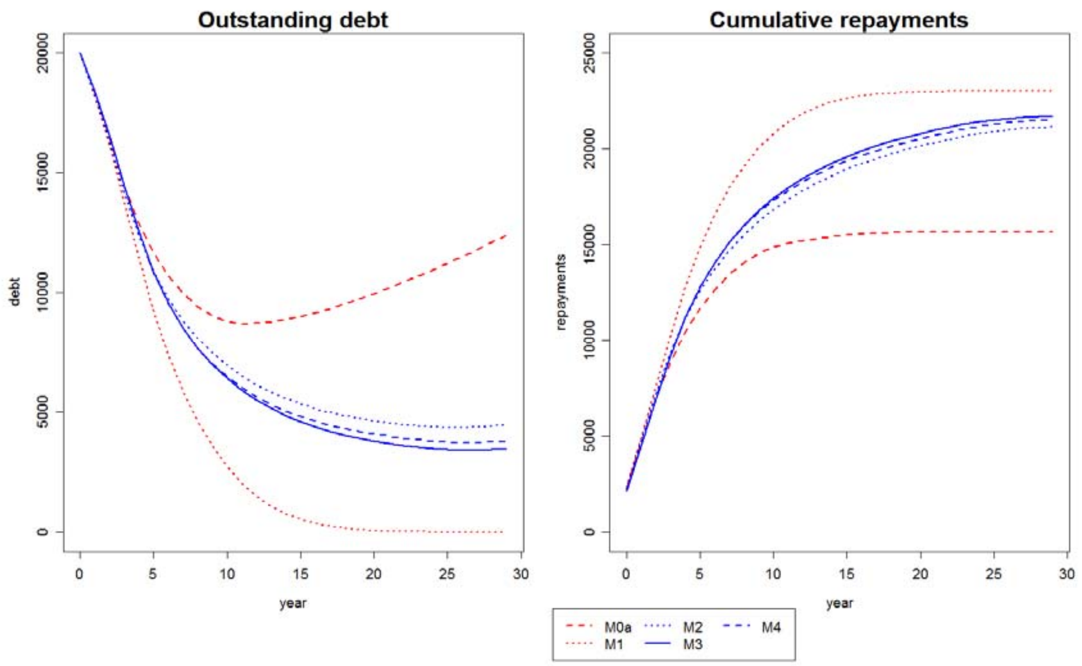
[28] The implication of labour force mobility can be explored by incorporating labour force transition models into the simulations. While not detailed here, in Higgins (2011) a number of dynamic labour force models are developed using HILDA data and incorporated into simulation of HECS repayments and costs.

$51,071-$53,754	5.0
$53,755-$57,782	5.5
$57,783-$62,579	6.0
$62,580-$65,873	6.5
$65,874-$72,492	7.0
$72,493-$77,247	7.5
$77,248 and above	8.0

Source: DEST, 2007

100 simulations were performed for each earnings model considered for all individuals within the subpopulation for the 30 year projection period. Results of the average calculations for 100 simulations are in Figure 7.

FIGURE 7 AVERAGE PROJECTED OUTSTANDING DEBT AND REPAYMENT ESTIMATES.



Model M.0a, a variation on the deterministic model M.0, is included to illustrate the implication of static earnings. Model M.0a assumes that all residual variation is due to static permanent individual differences. As such, M.0a allows for unobserved individual variation, but ignores the possibility of earnings mobility. The implication is seen in Figure 7; between 10 and 15 years, debtors above the minimum threshold have repaid, while those under the threshold will remain under, and their debt continues to grow with accrued interest. In contrast, under stochastic earnings models, earnings mobility enables movement across the minimum threshold. Under the naive stochastic model M.1, all variation is transitory, and the extreme movements in wage are such that all debtors frequently move above the minimum threshold, leading to complete debt repayment. Under models M.2, M.3 and M.4 a combination of permanent differences and transitory variation yields results between the extremes of models M.0a and M1. Although mobility under the stochastic models will result in some full-time individuals falling below the minimum threshold, the same mobility leads to an increase in repayments and reduction in debt for others. The simulations suggest that, on balance, this mobility increases the average repayment compared to static earnings assumptions. It is also apparent that clear differences in debt and repayments between M.2, M.3 and M.4 do not manifest until at least 5 years into the projection period. This highlights the importance of using long data periods for model assessment.

7. Conclusion

The mobility in earnings within static labour force states across the first seven waves of HILDA is clear. A review of the literature, coupled with exploratory analysis of the residuals from a mean fit, justified a variance components approach to modelling. A model with a static permanent component was compared with models that allowed for permanent shocks. Non-normal errors and serial correlation in the transitory component were also considered. An assessment of short-term predictions yielded no clear preferred model, yet theory and empirical investigation of the residual covariance structure suggests that permanent shocks through a random walk may be appropriate for full-time workers. Illustrations of model variability applied to HECS debt highlights the importance of allowing for realistic earnings mobility through variance decomposition. The simulations also show the importance of using long data periods for model assessment. As there is insufficient longitudinal HILDA data to separately allow model training and adequate assessment of predictions, an alternative is to use cross-validation. This will be applied to the HILDA data for both model development and testing in the future.

Clearly, work remains to explore the existing model processes further. For example, the variance estimates of the transitory and permanent shocks for M.3 could be permitted to differ with respect to age for predictions. Further, evidence was given that the variance components of hourly wage may differ by duration at the same employment state. The sensitivity of the parameter estimates to additional data should also be examined further. Waves 8 and 9 of HILDA will be added to the data and used to refit the models. Notably, two additional waves of data reduces the variability in the estimates for long-lagged earnings shocks and may yield different parameter estimates (e.g., using the first seven waves enabled only two data points for estimating $r_{i,t+5} - r_{i,t}$, whereas this doubles to four points when Waves 8 and 9 are included).

It is also recognised that alternative model processes need to be considered. There is considerable scope for deeper exploration of wage variation when individuals change their employment state between part and full-time. Heterogeneity in the error processes across individuals could be considered and the existing processes should be scrutinised further. For example, while a random walk for the permanent component replicates the observed increase in variance with age and improves the predictions, the process fit implies small annual changes. While this may be appropriate for some permanent shocks (e.g., wage increases above or below AWOTE), other shocks may be greater in magnitude, yet less frequent (e.g., change of occupation). Exploring how changes in hourly wage are correlated with changes in employment characteristics will enable a better understanding of the drivers of wage variation and may lead to improved dynamic wage models.

Appendix A1 Estimating the variance components for earnings model M.2.

First define the unobserved earnings shock as the first difference of the residuals from the mean fit:

$$g_{i,t+1} = \Delta r_{i,t+1} = r_{i,t+1} - r_{it} = \Delta u_{i,t+1} + \Delta \upsilon_{i,t+1} = w_{i,t+1} + \Delta \upsilon_{i,t+1}.$$

The variance is:

$$\operatorname{var}\left(g_{i,t+1}\right) = \operatorname{var}\left(w_{i,t+1} + \Delta \upsilon_{i,t+1}\right) = \sigma_w^2 + 2\sigma_\upsilon^2 - 2\operatorname{cov}(\upsilon_{it+1}, \upsilon_{it}),$$

where it is assumed that w and υ are orthogonal. Extending this to greater lags:

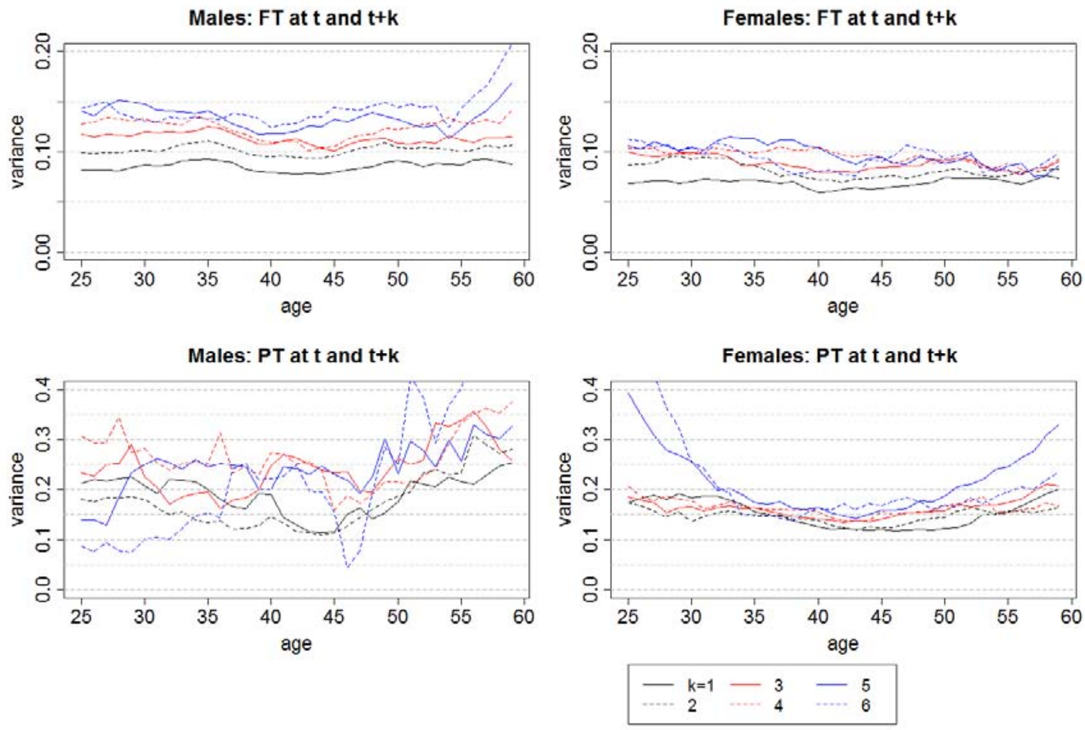
$$r_{i,t+k} - r_{it} = u_{i,t+k} - u_{it} + \upsilon_{i,t+k} - \upsilon_{it}, \qquad \text{and}$$

$$\operatorname{var}\left(r_{i,t+k} - r_{it}\right) = \operatorname{var}\left(\sum_{s=t}^{t+k-1} w_{is} + \upsilon_{i,t+k} - \upsilon_{it}\right) = k\sigma_w^2 + 2\sigma_\upsilon^2 - 2\operatorname{cov}(\upsilon_{i,t+k}, \upsilon_{it}),$$

since $u_{i,t+k} = u_{it} + \sum_{s=t}^{t+k-1} w_{is}$.

For the analysis the residuals from the regressions (i.e., $\hat{r}_{it}$ and $\hat{g}_{it}$) are used in place of r_{it} and g_{it}. I consider separate variance decomposition models for part-time and full-time earners. While it is assumed that the same residual functional form (M.2) applies to both part-time and full-time workers, it is clear from the variances of the earnings innovations in Figure A1 that there is considerably more variability in log hourly wage among part-time earners, and as such parameter estimates will differ.

FIGURE A1 VARIANCE OF MALE AND FEMALE EARNINGS INNOVATIONS ($\hat{r}_{i,t+k} - \hat{r}_{it}$) FOR ALL YEARS, 2001 TO 2007.



There is greater variance in changes in the earnings equation residuals for larger time gaps (for different k). This is expected if there are permanent shocks that have a multiplicative effect over time and is further support for a random walk for the permanent component.[29] In the first instance under the simplifying assumption that there is no serial correlation in the transitory shocks, then:

$$\operatorname{var}\left(r_{i,t+k}-r_{it}\right)=k\sigma_w^2+2\sigma_\upsilon^2 .$$

Therefore, assuming at least three years of panel data, enabling calculations of $\operatorname{var}\left(r_{i,t+1}-r_{it}\right)$ and $\operatorname{var}\left(r_{i,t+2}-r_{it}\right)$, a solution for σ_w^2 and σ_υ^2 can be found.

Conveniently, this approach is still robust if the transitory process is MA(q) and there are at least q+3 time periods. For example, if the transitory shock follows an MA(1) process, then $\operatorname{cov}(\upsilon_{i,t+k},\upsilon_{it})=0$, for $k\geq 2$, and the relations $\operatorname{var}\left(r_{i,t+2}-r_{it}\right)=2\sigma_w^2+2\sigma_\upsilon^2$, and $\operatorname{var}\left(r_{i,t+3}-r_{it}\right)=3\sigma_w^2+2\sigma_\upsilon^2$ can be used to solve for the variance of the shocks.

I first proceed under the assumption that the process in the transitory component is no greater than an MA(2) and use $\operatorname{var}\left(r_{i,t+k}-r_{it}\right)$ for k = 3, 4, 5 and 6 to find estimates for the variances of the shocks. For both sexes the estimates vary according to the age range considered as seen in Figure A1. Results are given in Table 6 in the paper for Model M.3.

For model M.2, the estimation process is even simpler, since we assume $\sigma_w^2=0$, so $\operatorname{var}(r_{i,t+k}-r_{it})=2\sigma_\upsilon^2$ assuming the process in the transitory component in no greater than an MA(2). Estimates for σ_υ^2 for M.2 are given in Table 5.

References

ABS (Australian Bureau of Statistics) (2010) *Average Weekly Earnings*, February, ABS Cat 6302.0, Canberra: AGPS

Alvarez,.J., Browning, M. and Ejrnaes, M. (2002) *Modelling Income Processes with lots of heterogeneity*, Unpublished Manuscript, Universidad de Alicante and University of Compenhagen.

Andreassen, L., Fredriksen, D. and Ljones, O. (1996) 'The Future Burden of Public Pension Benefits: A Microsimulation Study', *Microsimulation and Public Policy*, A.Harding (ed.), Elsevier, pp329-357.

Baker, M. (1997) 'Growth-rate heterogeneity and the covariance structure of life-cycle earnings', *Journal of Labor Economics*, 15(2):338-75.

[29] Notably this pattern is almost identical to the variances of the differences between log(wage) over various lags due to the low explanatory power of the observed covariates.

Baker, M., Solon, G. (2003) 'Earnings Dynamics and Inequality among Canadian men, 1976-1992: Evidence from longitudinal income tax records', *Journal of Labour Economics*, 21(2):289-321.

Becker, G.S. (1962) *Human Capital Theory*, Columbia University Press, New York.

Breusch, T. and Mitchell, D. (2003) *Australian Family Income Dynamics: Preliminary Evidence from the NLC Project, 1997 and 2000*, Negotiating the Life Course Discussion Paper Series, DP-014, April 2003

Chamberlain, G., Hirano, K. (1999) 'Predictive distributions based on longitudinal earnings data', *Annales d'Economie et de Statistique*, 55-56:211-242.

DEST (Department of Education, Science and Training) (2007) *Fee-Help Information 2007*, Canberra: AGPS.

Dickens, R. (2000) 'The evolution of individual male earnings in Great Britain: 1975-95', *The Economic Journal*, 110:27-49.

Geweke, J. and Keane, M. (2000) 'An empirical analysis of income dynamics among men in the PSID: 1968-1989', *Journal of Econometrics*, 96:293-356.

Harris, A. and Sabelhaus, J. (2003) *Projecting Longitudinal Earnings Patterns for Long-Run Policy Analysis*, Technical Paper Series, Congressional Budget Office, Washington DC

Higgins, T. (2011), Essays in the Development and Costing of Income Contingent Loans, PhD thesis, Australian National University.

Hirano, K. (2002) 'Semiparametric Bayesian Inference in autoregressive panel data models', *Econometrica*, 70(2):781-799.

Jenkins, S. (2000) Modelling household income dynamics, *Journal of Population Economics*, 13:529-567.

Kalb, G. and Scutella, R. (2002) *Estimation of Wage Equations in Australia: Allowing for Censored Observations of Labour Supply*, Melbourne Institute Working Paper No. 8/02

Keegan, M., Thurecht, L. (2008) *APPSIM – Modelling of Earnings*, APPSIM Working Paper No. 10, NATSEM

Lettau, M. (1994) *Compensation in Part-Time Jobs versus Full-Time Jobs: What if the Job is the Same?*, BLS Working Paper 260, U.S. Department of Labor, Bureau of Labor Statistics.

Lillard, L. and Willis, R. (1978) Dynamic aspects of earnings mobility. *Econometrica*, 46:985-1012.

MaCurdy, T. (1982) ‘The use of time series processes to model the error structure of earnings in a longitudinal data analysis’, *Journal of Econometrics*, 18:83-114.

Meghir, C. and Pistaferri, L. (2004) ‘Income Variance Dynamics and Heterogeneity’, *Econometrica*, 72(1):1-32.

Moffitt, R.A., and Gottschalk, P. (2002) ‘Trends in the transitory variance of earnings in the United States’, *The Economic Journal*, 112:C68-C73.

O'Donoghue, C. (2001) ‘Dynamic Microsimulation: A Methodological Survey’, *Brazilian Electronic Journal of Economics*, Department of Economics, Universidade Federal de Pernambuco, vol. 4(2), December.

Ramos, X. (2003) ‘The covariance structure of earnings in Great Britain, 1991-1999’, *Economica*, 70:353-374.

RBA (Reserve Bank of Australia) (2009a) ‘G06 Labour costs’, *Statistics*, Sydney.

____(2009b) ‘G2 Consumer Price Index’, *Statistics*, Sydney.

Rodgers, J.R (2004) Hourly wages of full-time and part-time employees in Australia, *Australian Journal of Labour Economics*, 7(2):231-254.

Swan, N. (1997) *Problems in Dynamic Modeling of Individual Incomes*, Swedish Conference on Microsimulation, Stockholm.

Watson, N. (2010) *HILDA User Manual – Release 8*, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.