Economics Working Paper Series

2015 - 06

Understanding Wage Inequality in Australia

Arpita Chatterjee, Aarti Singh
& Tahlee Stone

March 2015

Understanding Wage Inequality in Australia

Arpita Chatterjee* **Aarti Singh†** **Tahlee Stone‡**

March 25, 2015§

Abstract

In this paper we document rise in wage inequality in Australia over the last decade. A key driver of this inequality is unobservable or residual inequality. We decompose residual wage inequality into permanent and transitory components. The estimates of the permanent shock from the first difference approach allow us to reconcile life-cycle wage inequality.

Keywords: Wage inequality, life-cycle, permanent and transitory wage shocks

JEL codes: E24, E25.

*University of New South Wales.

†University fo Sydney. Corresponding author Email: aarti.singh@sydney.edu.au

‡University of New South Wales.

§This paper builds on Tahlee Stone' honours thesis. We thank Stanley Cho, Denzil Fiebig, Lisa Magnani, James Morley and Yasaman Naghiloo for helpful comments on an earlier draft.

1 Introduction

Growth in income has been accompanied by rising inequality in Australia. This trend has also been observed for other advanced economies including the United States, Canada and the United Kingdom.[1] For Australia, the growth in inequality is particularly pronounced in the last decade. The objective of this paper is to characterize wage inequality in Australia, across time and over an individual's life-cycle, using the Household, Income and Labour Dynamics in Australia (HILDA) data.

Recent empirical studies document rising trends in income and wage inequality in Australia. For example, Greenville, Pobke and Rogers (2013), Wilkins (2013) and Whiteford (2013) analyse inequality trends in Australia using ABS cross-sectional data. However, to our knowledge, an analysis of life-cycle wage inequality and the decomposition of wage shocks into permanent and transitory components has not been undertaken for the Australian economy.[2] In this paper we fill this gap in the literature by using the panel structure of HILDA data to examine life-cycle inequality and estimate permanent and transitory components of wage risk. Reliable estimates of permanent wage risk, which is primarily uninsurable, are often needed to understand the macroeconomic implications of market incompleteness.[3]

To empirically distinguish permanent and transitory wage shocks in the Panel Study of Income Dynamics (PSID) data for the US, Gottschalk and Moffit (1994) were the first to recognize the relevance of an unobserved components model. Using an ARMA(p,q) model for residual wage, they find that the permanent component has a unit root while the transitory

[1]See Heathecote, Perri and Violante (2010) for the U.S.; Brzozowski, Gervais, Klien and Suzuki (2010) for Canada and Blundell and Etheridge (2010) for the UK.

[2]An exception to this is Barrett, Crossley and Worswick (2000). They argue that a substantial portion of overall income inequality was driven by transitory income shocks because the rise in consumption inequality was lower than the rise in income inequality during 1975-1993. They use consumption and income data from the Australian Bureau of Statistics (ABS) Household Expenditure Survey (HES).

[3]Krebs (2003), Pijoan-Maas (2006), Singh (2010) among others argue that the macroeconomic effects of uninsurable income risk can be large.

shocks are uncorrelated. This simple specification of the wage process has gained particular traction among macroeconomists who study the quantitative effects of market incompleteness.

Heathecote, Perri and Violante (2010) examine income dynamics in the United States using biannual PSID data (1967-2005). Their model, like that of Gottschalk and Moffit (1994), features a permanent component with a unit root and a serially uncorrelated transitory component. Heathcote et al. (2010) draw upon both the labour economics and macroeconomic traditions and estimate two versions of the permanent-transitory model on wage residuals. Following the labour economics literature (for example, Abowd and Card, 1989, Meghir and Pistaferri, 2004 and Blundell, Pistaferri and Preston, 2008), they estimate the variance of wage shocks using moment conditions based on log residual wage growth. They also use moment conditions based on log wage level following the macroeconomics literature (for example, Storesletten, Telmer and Yaron, 2004, Guvenen, 2007 among others). Surprisingly, the estimates are strikingly different if one uses the first difference or the level approach. This *empirical puzzle* is also reported in other advanced economies. For example, Brzozowski, Gervais, Klien and Suzuki (2010) report it for Canada; Fuchs-Schundeln, Krueger and Sommer (2010) for Germany and Floden and Domeij (2010) for Sweden. In this paper we show that this empirical puzzle is also seen in the Australian data. Overall, this suggests that the conventional permanent-transitory model is potentially mis-specified. However, identifying the nature of this mis-specification is left for future research.

Using the estimates of the permanent and transitory wage shocks we show that while the estimates from either method can generate a rise in inequality over time, it is the estimates from the first difference approach that are better suited to describe life-cycle wage inequality in Australia.

The rest of the paper is structured as follows. Section 2 provides a brief overview of the data. Section 3 documents inequality across time and over an individual's life-cycle. Section 4 uses the panel dimension of HILDA to decompose residual wage into permanent and transitory components using level and first difference approach. The estimates of wage risk are different using the two approaches. Section 5 concludes.

2 Data and measurement

In this section we describe the data and discuss sample selection. We also establish why we focus on hourly wage–a key component of labour income–in subsequent analysis.

2.1 Data and sample selection

We use individual-level data from HILDA, a longitudinal survey that draws a representative sample from private households living in Australia. The survey began in 2001 and has eleven waves available to date. In the survey, all sample household members aged 15 years and over are interviewed on an annual basis. The first wave in 2001 interviewed 6,872 households and 13,969 individual respondents. In 2011, 2,153 households and 5,477 individuals were included in the survey. Our initial unbalanced dataset consists of 26,028 individuals and 147,823 observations. To stay consistent with the existing literature that studies wage dynamics in other advanced countries, our sample consists of male full-time workers between age 23 and 60. We exclude self-employed workers and individuals with missing demographic and income information. The final sample therefore includes 3,906 individuals and 23,261 observations.[4]

In HILDA most income variables, except weekly wage used in the construction of hourly wage, are reported in retrospective terms as the amount earned in the previous financial year. These income measures are converted into real terms using the Consumer Price Index.[5] We construct the real hourly wage rate as the ratio of earnings to the usual hours worked per week. The earnings are weekly gross wages or salaries collected from

[4]See Appendix A for a detailed outline of the sample selection procedure.

[5]The Consumer Price Index is obtained from the Statistical Table G1 of the Reserve Bank of Australia (RBA).

TABLE 1: Descriptive Statistics for Full-time Male Sample

Variable	2000-2001	2010-2011
Age (years)	39.4	41.4
% Graduates or higher	27.7	29.6
% Married	65.0	58.8
% Union	37.9	31.4
% English speaking migrant	14.8	10.9
% Non-english speaking migrant	9.1	7.9
% Australian born	76.1	81.2
% Indigenous	1.0	1.5
Average hours worked	2323	2290
Average annual earnings (AUD)	70996	84714
Number of observations	1856	2268

all jobs and received in the current financial year and the hours worked per week variable is either the reported hours per week, or if individuals indicate that their hours vary, it is the average hours per week. We also construct a binary education variable where individuals with at least 15 or more years of education are classified as 'high' and those with less than 15 are classified as 'low'. Table 1 reports the key demographic characteristics of our sample.

2.2 Labour income versus hourly wage

Figure 1 plots the time series of average labour income and average pre-tax income, where pre-tax income is the annual gross weekly wages and salaries from the previous year. Both pre-tax income and labour income follow similar trends except for a noticeable rise in pre-tax income in 2009-10.[6] From the figure it is evident that labour income accounts for approximately 75% of the pre-tax income. As a result, the inequality patterns identified in our paper using the labour component of income are also likely to be a key

[6]This can potentially be due to public transfers in the form of government stimulus bonus payments following the Global Financial Crisis.

determinant of aggregate income inequality trends for Australia.

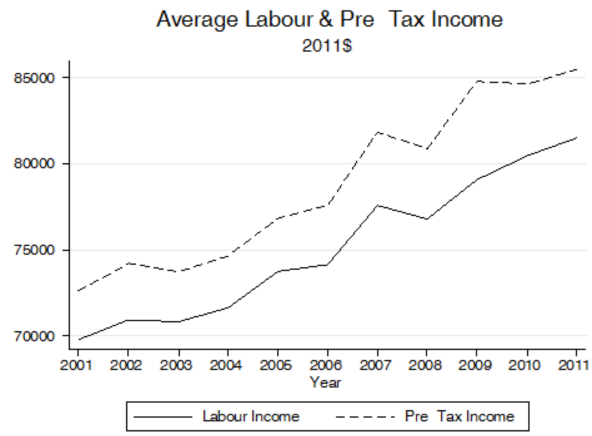

Figure 1: Trends in average labour income and pre-tax income.

Changes in labour income over time could be either due to changes in the hourly wage or changes in the number of hours worked or both. Figure 2 (left and right panel) plots trends in real average hourly wage and weekly hours for full-time working males. While the average real hourly wage has increased substantially, by approximately 20%, over the sample period, the trend is hours has been quite stable. Together, these two plots suggest that the average full-time hourly wage has driven the growth in labour income, a conclusion also reached in Greenville et al. (2013).

Figure 2: Trends in the two key components of labour income.

Growth in average real hourly wage is an indicator of the rising average standard of living in Australia. However, it is unclear whether this is accompanied by an increase in inequality. We examine the inequality trends in Australia in the following section.

3 Trends in inequality

In this section we document trends in inequality, both over time and across age.

3.1 Inequality over time

Has inequality changed over time in Australia? Figure 3 plots commonly used measures of wage inequality, variance of log hourly wage (left panel) and the Gini coefficient (right panel). Both measures exhibit a similar pattern of a substantial rise in inequality over the last decade in Australia, a fact documented in other studies. For example, our results suggest that the Gini coefficient grew 8.36% over the sample period; Greenville et al. (2013) find a 10.2% rise in the Gini coefficient of male hourly wage from 1998–2010 using ABS HES Survey which includes all individuals as opposed to only those who are of working age.

Figure 3: Inequality over time.

3.1.1 Observable versus unobservable

To identify sources of wage inequality we investigate the role of education and experience, two key observable characteristics, in explaining between-group inequality. The education wage premium, on average, has been stable over the sample. In the left panel in Figure 4 we plot the education wage differential, a ratio of the average wage of high to low education groups. Similarly, the experience wage premium has also not changed significantly (the right panel in Figure 4) over the sample period. The experience wage premium is calculated as the ratio of average wage of males between ages 44-55 to the average wage of males between ages 25-35. Since wage premium due to education and experience has remained stable over the sample period, we turn to consider within-group or residual inequality to examine sources of wage inequality.

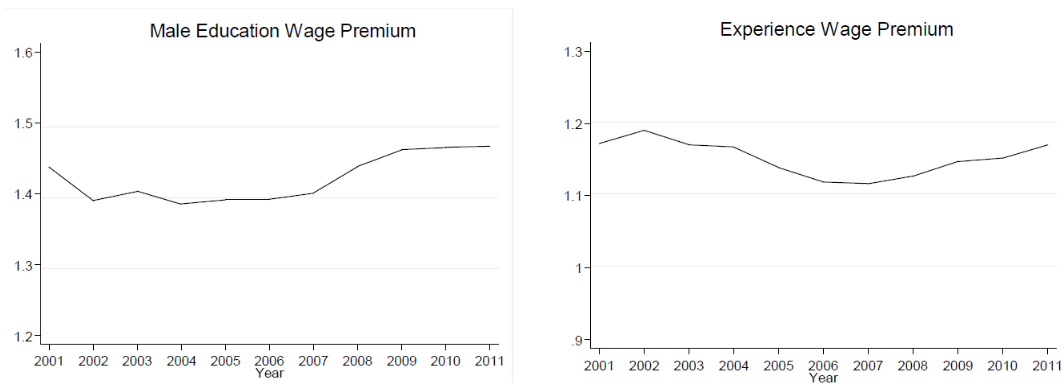

Figure 4: Education, experience and wage inequality.

The residual component of hourly wage, u_{it}, is given by:

$$y_{i,t} = X_{i,t}\varphi_{i,t} + u_{it}, \tag{1}$$

where $y_{i,t}$ is the log of hourly wage and $X_{i,t}$ is a vector of controls such as a cubic polynomial of age, education, migrant, married, indigenous, union member, health as well as state fixed effects.[7] Figure 5 plots the time-varying variance of $\hat{u}_{i,t}$ and the variance of log hourly wage. The residual wage inequality closely follows the "raw" inequality of log wage, as seen in Figure 3 left panel, and accounts for 81% of overall cross-sectional inequality.[8] This provides strong evidence that unobservable characteristics or residual inequality is an important driver of wage inequality in Australia. Is the residual inequality being driven by permanent or transitory

[7]This specification is similar to the set of controls used by previous studies such as Heathcote et al. (2010) and Huggett, Ventura and Yaron (2011). It has been adapted to the Australia economy by adding additional controls such as indigenous status, union member, health and state fixed effects. Including these additional controls, however, does not impact our findings. As the regressions are carried out by year, we do not control for time.

[8]Heathcote et al. (2010) also find that the residual inequality accounts for approximately 79% of overall wage inequality in the U.S data.

shocks to wages? Section 4 examines this in more detail.

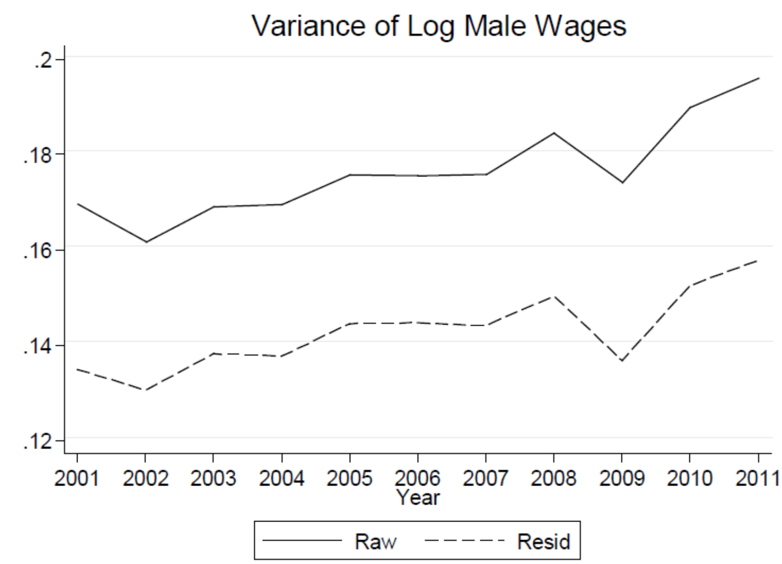

Figure 5: Residual wage inequality.

3.1.2 Life-cycle inequality

Following Deaton and Paxson (1994), we identify inequality patterns over an individual's life-cycle. To ensure that there are sufficient observations in each age-year bin we construct a rotating age panel. These centered 5 age-year cells are constructed by defining an individual's age as 'a' if their actual age lies within the range 'a-2' and 'a+2'. This ensures that there are at least 100 observations in each age-year bin.

In this strand of the literature, it is commonly understood that obtaining an accurate measure of life-cycle inequality which is entirely due to age effects is challenging due to the collinearity given by

$$Age(a) = Year(t) - Cohort(c). \tag{2}$$

Cohort effects are conditions specific to an individual's birth year while time effects capture changes in the aggregate economic environment in a particular year. Separating out pure age effects in the presence of these time and cohort effects is not feasible due to their inter-dependency. As a result, in constructing the age effects, we either control for time or cohort effects.[9]

3.1.3 Model specification and identification

To characterize the distribution across age, we report statistics measuring the mean and dispersion of wages using the following model:

$$h_{a,t} = \beta_0 + \beta_a D_a + \beta_c D_c + +\beta_t D_t + \epsilon_{a,t} \tag{3}$$

where β_a, β_c, and β_t represent the coefficients on age D_a, cohort D_c and year dummies D_t respectively. Because of the colinearity, we control for cohort (time) effects by restricting the year (cohort) effects to zero. The age-profile for each moment, mean and variance, is computed using the estimated vector of age coefficients, $\hat{\beta}_a$.

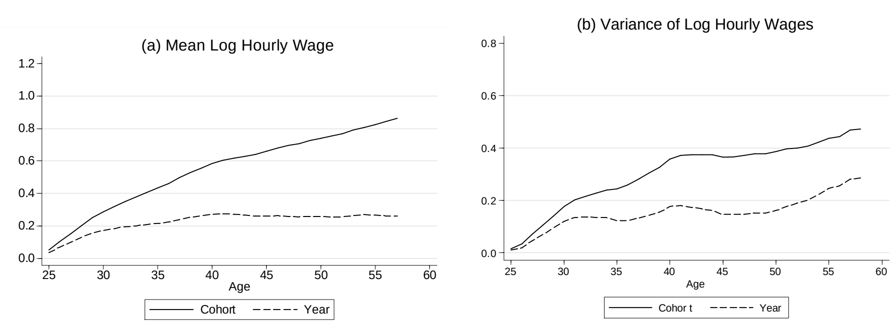

Figure 6: Mean and variance of log hourly wage.

Figure 6 plots the distribution of wags over the life-cycle. While controlling for cohort and time effects produces qualitatively similar results, the precise magnitudes are different depending on whether we control for

[9]See Heathcote, Storesletten, and Violante (2005) for discussion of this issue.

time or cohort effects. Controlling for cohort effects results in steeper profiles than controlling for time effects across all statistics. Similar findings have been reported in other related studies (for example, Heathcote et al., 2005, Brzozowski et al., 2010 and Heathcote et al., 2010). Heathcote et al. (2005) argue that this is due to the fact that the cohort approach does not control for the rising time trend. This is potentially the case for the Australian data as well. The rising time trend observed for the first and second moment of real hourly wage (left panels in Figure 2 and 3) is confounding with age effects and attributing to the steeper 'cohort' profiles observed in Figure 6.

Focusing on the specification that controls for time effects, the life-cycle pattern for the mean of real hourly wage exhibits the typical hump-shaped profile. In addition, we also notice that the wage inequality increases over the life-cycle. Note that inequality rises until around age 43-47 after which wage dispersion does not grow as rapidly. This is in contrast to studies such as Brzozowski et al. (2010) and Blundell and Etheridge (2010) that find evidence for a monotonic linear profile of the variance of wage. Our results are similar to the results in Heathcote et al. (2010) where the life-cycle inequality profile of log wage is non-linear. Overall, it is evident from the right panel of Figure 6 that wage inequality fans over the life-cycle. This lends support to the view that individuals encounter persistent income shocks over their working life. To investigate this further, we estimate the variance of permanent and transitory wage shocks in the following section and look closely into the sources of residual inequality, an important driver of wage inequality in Australia.

4 Dynamics of inequality

So far we have documented that inequality has increased over time and unobservables are the key driver of wage inequality in Australia, like many other advanced countries. In addition, we also document that wage inequality increases with age suggesting that wage shocks over an individual's life-cycle could be an important source of life-cycle wage inequality. Therefore, in this section we estimate these wage shocks. We decompose residual (within-group) wage into permanent and transitory components using an unobserved components model.

Following the literature, we specify a parsimonious permanent and transitory component model for the residual wage given in equation (1). Let z_{it} be the permanent component which follows a random walk such that

$$\begin{aligned} \widehat{u}_{it} &= z_{it} + \epsilon_{it} \\ z_{it} &= z_{it-1} + \eta_{it} \end{aligned} \tag{4}$$

where η_{it} and ϵ_{it} are innovations to the permanent and transitory component respectively. The shocks are uncorrelated over time, i.i.d. across individuals, and orthogonal to each other. Let $\sigma^2_{\eta t}$ and $\sigma^2_{\epsilon t}$ and denote the variances of the two shocks at date t.

To identify the stochastic volatility of these shocks, we derive the moment conditions using the growth rate of the residuals and their levels. Rewriting equation (4), we get

$$\widehat{u}_{it} = z_{it-1} + \eta_{it} + \epsilon_{it} \tag{5}$$

$$\widehat{u}_{it+1} = z_{it} + \eta_{it+1} + \epsilon_{it+1} = z_{it-1} + \eta_{it} + \eta_{it+1} + \epsilon_{_{it+1}} \tag{6}$$

To write the moment condition using growth rates, let

$$\Delta\widehat{u}_{i,t} = \widehat{u}_{i,t} - \widehat{u}_{i,t-1} = \eta_{it} + \epsilon_{it} - \epsilon_{it-1} \tag{7}$$

and

$$\Delta\widehat{u}_{i,t+1} = \widehat{u}_{i,t+1} - \widehat{u}_{i,t} = \eta_{it+1} + \epsilon_{it+1} - \epsilon_{it}. \tag{8}$$

Then

$$cov(\Delta\widehat{u}_{i,t+1}, \Delta\widehat{u}_{i,t}) = -cov(\epsilon_{it}, \epsilon_{it}) = -\sigma^2_{\epsilon t} \tag{9}$$

$$var(\Delta\widehat{u}_{i,t}) = \sigma^2_{\eta t} + \sigma^2_{\epsilon t} + \sigma^2_{\epsilon t-1}. \tag{10}$$

The following moment conditions use level data

$$var(\widehat{u}_{it}) - cov(\widehat{u}_{it+1}, \widehat{u}_{it}) = \sigma^2_{\epsilon t} \tag{11}$$

$$var(\widehat{u}_{it}) - cov(\widehat{u}_{it}, \widehat{u}_{it-1}) = \sigma^2_{\eta t} + \sigma^2_{\epsilon t} \quad (12)$$

Using equations (9) and (10) we estimate the variance of the two shocks based on the moment conditions from the first difference approach. Similarly, equations (11) and (12) provide an alternative method for estimating the variances of the two shocks by using the level data.

4.1 Results

Using the moment conditions, Figure 7 plots the variances of the permanent and transitory shocks to wages. Our estimates suggest that the variance of transitory shock is larger than the variance of the permanent shock irrespective of whether we use level or difference approach. This is consistent with the finding in Barrett et al. (2000). But since the impact of a transitory shock does not persist, it is the effect of the permanent shock that could potentially explain the fanning out of the wage distribution over the life-cycle.

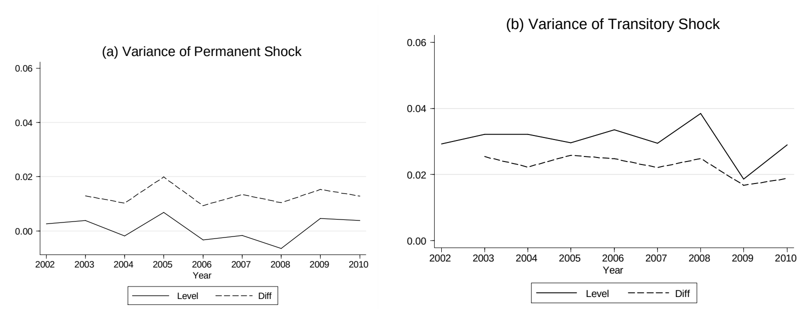

Figure 7: Estimates of variances of permanent and transitory shocks to residual wage.

Even though the estimated variance of the permanent shock follows a similar pattern over the sample period using both approaches, the magnitude of the estimates is different. Using the first difference approach, the variance of the permanent shock is higher than the variance from the levels approach. Unless there is model mis-specification, the two approaches

TABLE 2: Estimates of Wage Dynamics: Cross-Country Comparison

Country	Data	Sample period	σ_η^2	
			Level	Difference
Australia	HILDA	2001-2011	.001	.013
U.S.	PSID	1967-2002	.019	.028
Canada	SLID	1993-2005	-	.055
Germany	GSOEP	1984-2004	.010	.030
Sweden	LINDA	1978-2004	.005	.040

should generate similar estimates. This inconsistency has been observed in related literature and Table 2 reports the estimates from other studies.

Each study uses the panel dimension of its data source and employs a similar model to estimate the shocks. Most of these studies acknowledge that the estimate of variance of the permanent shock from the first difference approach is too large to explain the life-cycle inequality in their respective countries. These studies, therefore, typically conclude that the level estimates of the permanent shock are better suited to explain life-cycle inequality. However, since the estimate of the variance of the permanent shock in Australia using the first difference approach is closer to the level estimate in U.S. and Germany, will it help in reconciling life-cycle wage inequality in Australia? We explore this further in the following subsection.

4.2 Evaluating the estimates

Although the estimates point towards model mis-specification, the unobserved components model used in our analysis is extensively used in the quantitative macroeconomics literature. Therefore, in this section we explore which sets of estimates from the unobserved components model come closer in matching wage inequality in the data.

Since residual wage inequality captures key trends in the inequality of log wage, we simulate the residuals using the estimates of the permanent and transitory shocks based on moment conditions that use either the growth rate or the level of residuals. We compare this with the actual data. To make the comparison more appropriate the actual data, which is an unbalanced panel, has been appropriately adjusted. We include only individuals in years for which we have estimates of variances of permanent and transitory shock (2003-2010).[10] We also exclude individuals who exit and re-enter the survey. The resulting sample is an alternative to using a balanced panel for this analysis. A balanced panel would be easier in many regards; however, it would reduce the sample size considerably and introduce potential attrition bias.[11]

Figure 8 plots the simulated residuals using the level and first difference estimates of wage risk (dash and dotted lines respectively) together with the actual age profile (solid line). The simulated residuals are plotted on the left vertical axis and the actual residuals are plotted right vertical axis. The figure shows that both estimates produce a similar trend in the residuals and this trend matches the data.[12]

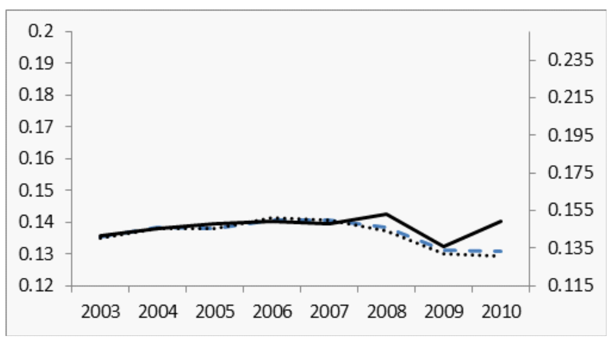

Figure 8: Actual versus simulated residuals.

In spite of the fact that the two measures produce similar trends in residual inequality, a simple calculation using the average value for the variance of the permanent shock suggests that the estimates using the first difference approach are better suited to explain the rise in inequality over the life-cycle. Over 32 years of working life from age 25-57, while the level estimate

[10]Data from 2002 is used to obtain the initial value of the permanent income for individuals in 2003.

[11]See Table A2 in the Appendix for more details.

[12]In instances where the estimate of the variance is negative, we have set it equal to zero in our simulations.

imply that the wage distribution would fan out by 0.03, the growth estimate suggest an increase of 0.42. In the data, however, wage inequality over the life-cycle increases by 0.3-0.5, see the right panel in Figure 6. Therefore, for Australia, unlike the U.S., Germany and other advanced economies that typically report this empirical puzzle, the growth estimates provide a better resolution of the life-cycle wage inequality.

4.3 Discussion

Blundell, et. al (2008) estimate the income risk using a model with correlated transitory shocks with $0 < \theta < 1$. As a result, the permanent-transitory specification of the residuals can be generalized and written as

$$\begin{aligned} \widehat{u}_{it} &= z_{it} + \epsilon_{it} + \theta\epsilon_{it-1} \\ z_{it} &= z_{it-1} + \eta_{it}. \end{aligned} \tag{13}$$

Then using the first difference approach, the moment condition that enables us to identify the variance of the permanent shock is given below

$$var(\Delta\widehat{u}_{i,t}) = \sigma^2_{\eta t} + \sigma^2_{\epsilon t} + (\theta-1)^2\sigma^2_{\epsilon t-1} + \theta^2\sigma^2_{\epsilon t-2}. \tag{14}$$

Therefore if the model is mis-specified with $\theta = 0$, using equation (12) to compute the variance of the permanent shock will result in a substantial upward bias even if θ is small. However, the moment conditions based on the level wage data to identify the variance parameters remain unchanged if we allow for serially correlated transitory shocks. To see this consider the following conditions

$$var(\widehat{u}_{it}) - cov(\widehat{u}_{it+1}, \widehat{u}_{it}) = (1-\theta)\sigma^2_{\epsilon t} + \theta^2\sigma^2_{\epsilon t-1} \tag{15}$$

$$var(\widehat{u}_{it}) - cov(\widehat{u}_{it}, \widehat{u}_{it-1}) = \sigma^2_{\eta t} + \sigma^2_{\epsilon t} + \theta^2\sigma^2_{\epsilon t-1} - \theta\sigma^2_{\epsilon t-1} \tag{16}$$

Hence, the level estimates are robust to this model mis-specification. See Domeij and Floden (2010) for a detailed discussion of this issue.

Since we don't have estimates of the serial correlation parameter, our conjecture is that such an estimate would potentially make the growth estimate of permanent risk even better suited to explain life-cycle inequality in Australia. However, the level estimate will continue to be too low to explain life-cycle wage inequality.

5 Conclusion

In characterizing wage inequality in Australia we document that income inequality has risen in Australia over the last decade, in line with the related literature. We further document that the key driver of this inequality is unobservable or residual inequality. Using well established tools in the empirical macroeconomics literature, we decompose residual wage inequality into permanent and transitory components. We establish that the estimates of the variances of the permanent and transitory shocks are very different depending on the moment conditions, in line with the international literature. For Australia, the estimates of the variance of the permanent shock from the first difference approach come closer in reconciling life-cycle wage inequality.

References

[1] Abowd, J. and Card, D. 1989. On the covariance structure of earnings and hours changes, Econometrica, 57, 411-45

[2] Barrett, G., Crossley, T. and Worswick, C. 2000. Consumption and income inequality in Australia, Economic Record, 76, 116–138

[3] Blundell, R., Pistaferri, L. and Preston, I. 2008. Consumption inequality and partial insurance. American Economic Review, 98, 1887-1921

[4] Blundell, R. and Etheridge, B., 2010. Consumption, income and earnings inequality in Britain. Review of Economic Dynamics, 13, 76-102

[5] Brzozowski, M., Gervais, M., Klein, P. and Suzuki, M. 2010 Consumption, income and wealth inequality in Canada. Review of Economic Dynamics, 13, 52-75

[6] Deaton, A., and Paxson, C. 1994. Intertemporal choice and inequality. Journal of Political Economy, 102, 437-67

[7] Floden, M. and Domeij, D. 2010. Inequality Trends in Sweden 1978-2004. Review of Economic Dynamics, 13, 179–208

[8] Fuchs-Schundeln, N., Krueger, D. and Sommer, M. 2010. Inequality trends for Germany in the last two decades: A tale of two countries, Review of Economic Dynamics, 13, 103-132

[9] Gottschalk, P., and Moffit, R. A. 1994. The growth of earnings instability in the US labor market. Brookings Papers on Economic Activity, 25, 217-272

[10] Greenville, J., Pobke, C. and Rogers, N. 2013, Trends in the distribution of income in Australia, Productivity Commission Staff Working Paper

[11] Heathcote , J., Storesletten, K., and Violante, G. L. 2005. Two Views of Inequality over the Life Cycle. Journal of the European Economic Association: Papers and Proceedings, 3, 765-75

[12] Heathcote, J., Perri, F., and Violante, G. L. 2010. Unequal we stand: An empirical analysis of economic inequality. Review of Economic Dynamics, 13, 15-51.

[13] Heathcote, J., Storesletten, K., and Violante, G. L. 2010. The macroeconomic implications of rising wage inequality in the United States. Journal of Political Economy, 118, 681-722

[14] Heathcote, J., Perri, F., and Violante, G. L. 2010. Unequal we stand: An empirical analysis of economic inequality. Review of Economic Dynamics, 13, 15-51.

[15] Huggett, M., Ventura, G., and Yaron, A. 2011. Sources of lifetime inequality. American Economic Review, 101, 2923-2954

[16] Krebs, T. 2003. Human capital risk and economic growth. Quarterly Journal of Economics, 118, 709-744

[17] Meghir, C., and Pistaferri, L. 2004. Income variance dynamics and heterogeneity. Econometrica, 72, 1-32

[18] Pijoan-Mas, J., 2006. Precautionary savings or working longer hours? Review of Economic Dynamics, 9, 326-352

[19] Singh, A. 2010. Human capital risk in life-cycle economies, Journal of Monetary Economics, 57, 729-738

[20] Storesletten, K., Telmer, C., and Yaron, A. 2004. Consumption and Risk Sharing Over the Life Cycle. Journal of Monetary Economics, 51, 609-33

[21] Whiteford, P. 2013. Australia: Inequality and Propserity Record, 74, 279-85

[22] Wilkins, R. 2013. Cross-sectional income distributions: How does HILDA compare with ABS income

A Data and sample selection

Like any micro-level household data source, HILDA survey has some commonly encountered issues that are briefly discussed below.

Imputation: Income variables taken from the HILDA survey have been treated for non-response using mainly a combination of Nearest Neighbor regression and Little and Su imputation methods. The rate of income non-response has decreased as the survey has become more established. In 2011, the percent of mean imputed income values was 3.4% and 4.8% for wages and salaries and total pre-tax income, respectively. In our analysis, we do not exclude imputed individuals as these figures represent relatively small portion of the sample.

Top-coding: Another data quality issue that can have implications for assessing inequality, particularly at the top of the wage distribution, is top-coding. This common censoring practice is used by surveys to protect the identification of top-income earners. The HILDA survey uses a weighted-mean top-coding procedure. They substitute any income values above a certain threshold with the weighted mean value of income for all individuals and households in the top-coded category. However this should have negligible implications for inequality measures as this procedure is carried on only 0.2%-0.4% of the sample.

Attrition: For HILDA, similar to other longitudinal studies, attrition over time has a cumulative effect in eroding the representativeness of the sample. Attrition will introduce bias to the survey if it is non-random; that is, if particular subgroups are more likely to attrit than others. In the HILDA survey individuals who are young, single, non-English speaking, unemployed or working in a low-skilled occupation have a higher likelihood of dropping out or refusing to participate in the study. These attributes are also associated with higher wage variability and individuals at the lower end of the wage distribution. Therefore, the higher propensity for attrition amongst people with these characteristics introduces a potential for bias within the sample towards excess stability in the wage distribution.

Sample representativeness: A limitation of HILDA is that migrants arriving in Australia after 2001 have little chance of entering the sample. This represents another source of declining representativeness relevant to this study. Australia's immigration policy is targeted at stable skilled migration. By not accounting for the inflow of skilled migrant into Australia, the HILDA survey may therefore underrepresent young, more highly educated human capital in the panel. This issue has the potential to understate inequality and introduce an age bias that may impact inequality measures taken over the life-cycle dimension.

TABLE A1: Sample selection

Description	Dropped	Remaining
Initial unbalanced sample		216 368
Missing wave information	15 026	201 342
Respondents only	53 519	147 823
Age restrictions	46 773	101 050
Employed-full time	47 444	53 606
Full-time student	339	53 267
Not self employed	9 300	43 697
Male only	16 318	27 649
Miscoded or zero weekly income	131	27 518
Miscoded or zero annual income	323	27 195
Drop if half the minimum wage	151	27 044
Missing demographic information	66	26 978
At least two consecutive waves	3 717	23 261

TABLE A2: Sample selection for simulation

Description	Dropped	Remaining
Sample based on Table A1		26 978
Keep Waves 2-10 only	7 993	18 985
Exclude those who exit and re-enter	2 507	16 478
At least three consecutive waves	1 065	15 413

Sample top-up: The HILDA survey aimed to minimize the effects of potential attrition bias and erosion of representativeness by replenishing the sample top-up in 2011. The main objective of administrators was to ensure the long-term representativeness of the panel. However, in the short-term the injection of new individuals and households had the potential to alter the structure of the wage distribution. However, Wilkins (2013) performed robustness checks on this sample top-up and found that the inclusion or exclusion of this additional sample has minimal impact on the 2010-2011 wage distribution.