Crawford School of Public Policy

TTPI

Tax and Transfer Policy Institute

Occupational mobility in the ALife data: how reliable are occupational patterns from administrative Australian tax records?

TTPI - Working Paper 5/2022
April 2022

Clara Hathorne
Tax and Transfer Policy Institute,
Crawford School of Public Policy, Australian National University

Robert Breunig
Tax and Transfer Policy Institute,
Crawford School of Public Policy, Australian National University

Abstract

The purpose of this paper is to compare the distribution of occupation and rates of occupational mobility in the ATO Longitudinal Information Files (ALife) and the Household Income and Labour Dynamics in Australia (HILDA) datasets. As tax is not occupation dependent, occupation data from tax records may not be reliable. We find that occupational mobility in the ALife data is less than half that in the nationally representative HILDA data. In contrast, the distribution of occupation and its relationship with most key socio-economic characteristics appear relatively similar across the two datasets. However, occupation evolves differently over time in the two datasets and there are some differences between sexes.

Keywords: Alife data, HILDA, occupational mobility

** We are grateful to Andrew Carter, Justin Holland, Son Nguyen and Hang Thi To for comments on an earlier draft. Son Nguyen provided exceptional assistance in producing the tables and graphs that examine occupational change by tax lodgement channel. We are grateful to Lucille Danks for providing the calculation of occupational change from the Labour Force Survey and the Census. We are grateful to Ben Jefferson for research support and his careful and clever graphics editing.*

Tax and Transfer Policy Institute

Crawford School of Public Policy

College of Asia and the Pacific

+61 2 6125 9318

tax.policy@anu.edu.au

The Australian National University

Canberra ACT 0200 Australia

www.anu.edu.au

The Tax and Transfer Policy Institute (TTPI) is an independent policy institute that was established in 2013 with seed funding from the federal government. It is supported by the Crawford School of Public Policy of the Australian National University.

TTPI contributes to public policy by improving understanding, building the evidence base, and promoting the study, discussion and debate of the economic and social impacts of the tax and transfer system.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Occupational Mobility in the ALife data: how reliable are occupational patterns from administrative Australian tax records?

Clara Hathorne and Robert Breunig*
Tax and Transfer Policy Institute
Crawford School of Public Policy
Australian National University

This version: 1 April 2022

Abstract:

The purpose of this paper is to compare the distribution of occupation and rates of occupational mobility in the ATO Longitudinal Information Files (ALife) and the Household Income and Labour Dynamics in Australia (HILDA) datasets. As tax is not occupation dependent, occupation data from tax records may not be reliable. We find that occupational mobility in the ALife data is less than half that in the nationally representative HILDA data. In contrast, the distribution of occupation and its relationship with most key socio-economic characteristics appear relatively similar across the two datasets. However, occupation evolves differently over time in the two datasets and there are some differences between sexes.

I. Introduction

In this paper, we examine the distribution of occupation and occupational mobility of the Australian labour force using two different data sources. The first of these is the widely used Household Income and Labour Dynamics in Australia (HILDA) survey. The second is a newly

* We are grateful to Andrew Carter, Justin Holland, Son Nguyen and Hang Thi To for comments on an earlier draft. Son Nguyen provided exceptional assistance in producing the tables and graphs that examine occupational change by tax lodgement channel. We are grateful to Lucille Danks for providing the calculation of occupational change from the Labour Force Survey and the Census. We are grateful to Ben Jefferson for research support and his careful and clever graphics editing.

available extract of administrative data taken from tax return records held by the Australian Taxation Office, the ATO Longitudinal Information Files (ALife). Our main purpose in undertaking this analysis is to assess the suitability of the occupation data in ALife for use in statistical and econometric analysis. Our focus is on occupation at the one-digit level, a classification which assigns workers into one of eight broad occupational categories.

As researchers increasingly use administrative tax records for research into questions around labour market and taxpayer behavior related to occupation, it is important to understand whether the occupational distribution and occupational change are captured with sufficient accuracy in the administrative tax records. The suitability of the ALife occupation data also depends upon the degree to which it captures the correlation between occupation and other demographic variables such as age, sex and income.

Occupation appears on the tax records of individuals, but it is not known how accurate the information is. Occupation is generally entered the first time an individual files a tax return, but, after that, the information is automatically pre-filled on subsequent tax returns for most individuals. Since tax liabilities are independent of occupation, there is little incentive for individuals to update the occupation field when they change occupations.

HILDA provides a benchmark for the distribution of occupation, occupational change and co-movements between occupation and key socio-demographic variables. As a nationally representative, longitudinal dataset which has been extensively benchmarked against other sources of national statistics, HILDA provides an excellent basis for comparison with ALife.

Our main finding is that occupation change is under-reported in the tax data; occupation change in ALife is less than half that in HILDA. This difference is relatively stable across time. In addition, rates of occupational mobility by sex are different in the two datasets. The overall distribution of occupation is roughly the same in the two datasets as are the broad trends in movements in the occupation distribution over the past two decades. However, the two datasets provide very different pictures of occupational mobility around the Global Financial Crisis where ALife shows much larger increases in occupational mobility than HILDA

Our analysis suggests that research using ALife which exploits cross-sectional variation in occupation by income, sex or age should, for the most part produce analysis that is representative of Australia. Analysis that relies on changes over time, such as fixed effects analysis, will suffer from reduced variation in the data and thus may be influenced by biases in which types of occupational changes are captured and which are not. If the unobserved factors which impact on reporting occupational change in ALife are correlated with unobservable which influence outcomes of interest, such analysis may be problematic.

II. Literature Review

Occupational and job mobility are of keen interest to researchers in labour economics due to the relationship between mobility, wages, and productivity. Occupational mobility provides evidence of a dynamic and functioning labour market, and, particularly for the young, is an important contributor to broader socio-economic mobility.

We undertake a very brief literature review to highlight a few of the issues surrounding occupational and job mobility and their relationships to economic outcomes. Studies in both Australia and the U.S. have found that occupational and job mobility are strongly linked to wage and productivity growth. There is also a link to wage inequality as mobility has allowed more qualified workers to pursue higher-paying jobs. Other research relates occupational mobility to demographic variables such as age, gender and education. We refer readers to the discussion and reference list in the papers cited below for further information.

Deutscher (2019) shows that a one percent increase in job-to-job mobility is associated with a one-half per cent increase in wage growth in Australia. Likewise, Kambourov and Manovskii (2009) find that there is a strong tie between occupation mobility and wage inequality in the US – finding that 90% of the increase in wage inequality in the 1990s is linked to an increase in occupational mobility. Sicherman (1990) analyses the effect of education on occupation mobility and finds that more educated workers have a higher degree of occupation-specific investment and thus have a lower degree of occupational mobility. However, when looking at a set of workers who start in the same occupation, higher educated workers have more upwards occupational mobility.

Many other papers have been dedicated to connecting occupational mobility to the business cycle; for example, Moscarini and Vella (2008) find that overall occupational mobility decreases with a workers' age, education, and family commitments. However, these effects are weaker (and sometimes reversed) when unemployment is high. This finding is consistent with the data we present below, revealing that following the 2008 global financial crisis occupational mobility in Australia jumped.

Lastly, a plethora of literature is dedicated to understanding intergenerational occupational mobility. In a World Bank paper, Sinha (2017) looks at cross-country disparities in intergenerational occupational mobility. They find that intergenerational occupational persistence is higher in poorer countries as a product of misallocation of talent. While these findings are not the focus of this paper, they demonstrate the importance of occupational mobility in understanding a range of micro- and macro-economic issues.

III. Data

For our analysis we utilize both the HILDA and ALife Datasets. The Household Income and Labour Dynamics in Australia (HILDA) data comes from an extensive survey of economic and social topics over the years of 2001 to 2018; see Watson and Wooden (2010). The other dataset we utilise for our analysis is the ALife dataset, the ATO's Longitudinal Information Files–see Abhayaratna, Carter and Johnson (2021). We use ALife 2017 which provides researchers with tax and superannuation (Australia's mandatory, individual contribution-based retirement scheme) data that range from 1991 to 2017. Occupation in HILDA is provided by respondents in an open-ended framework and then converted to standard code frames – see Watson and Summerfield (2009). This used to be the case with tax data, where tax filers would write a description of their occupation on their paper tax form and this would be transformed to standard occupation codes within the Australian Taxation Office.[1] In more recent years, the online system through which many people file includes a drop-down menu of occupations that tax filers can choose from. This is still subject to misreporting for two reasons. One is that this field is pre-filled from previous years, so may not capture occupational changes. The other reason is that the drop-down menu is quite limited, and people may mis-classify their occupation if they do not find a good fit for what they actually do in the menu.

These two datasets are quite different. HILDA is a household-based survey, originally based on a frame that covered all private dwellings in non-remote Australia in 2001. It was topped up in 2011, motivated primarily by a desire to improve the coverage of the data with respect to recent migrants to Australia. It includes rich socio-demographic data and a wide range of self-reported data on health, well-being, financial affairs, etc. It is subject to the usual reporting

[1] The ALife data were built from data in the ATO warehouse that contained both occupation codes and descriptions. Codes were derived from the descriptions where the code was missing.

errors found in all survey data. ALife is an extract of administrative tax records from the Australian Taxation Office (ATO). It has very accurate and detailed information on any information that taxpayers are required to provide to the ATO to determine their tax liability. It includes information provided by employers, banks, superannuation funds, financial service providers and self-reported data from taxpayers.

In what follows, we present snapshots from both data sources. We treat HILDA as providing an approximately "correct" picture of occupation in Australia, as it is benchmarked to national statistics and provides weights which allow for unbiased population-level cross-sectional and dynamic estimates of statistics. We use the weights for all of the descriptive statistics, but not for the figures. Weighted and unweighted figures (and descriptive statistics) are nearly identical. As occupation is not required by tax authorities to assess tax liability, we view the occupation data in ALife as less reliable. For other data items, relative reliability across the two datasets is difficult to assess and beyond the scope of this paper.

In looking at the distribution of occupation, our main variable of interest is a one-digit occupation classification code using the Australian and New Zealand Standard Classification of Occupations (ANZSCO 2006)—see Australian Bureau of Statistics (2006). This variable includes eight classifications of occupation that are consistent across the two datasets—professionals; managers; technicians and trade workers; community and personal service work; sales workers; clerical and administrative workers; machinery operators and drivers; and labourers. While this classification of occupation type is extremely simplified, it provides a common classification scheme across the two datasets which allows us to graphically examine how the distribution of occupation varies over time and across demographic groups.[2]

[2] Using two-digit or four-digit occupation codes results in very small samples in the HILDA data and consequent volatility and unreliability of estimates.

Occupation in ALife

Using the occupation data in ALife requires some care. A first issue of which researchers should be aware is that the occupational classifications used by the ATO have changed three times in the past two decades. From 1991-2002, the Australian Standard Classification of Occupations (ASCO), first edition was used. This was replaced by the second edition ASCO from 2003 – 2008 and then ANZSCO 2006[3] for the period from 2009 onwards.

In the interest of making all possible information available to researchers, the ALife data include four occupation variables. The first two variables are the two versions of ASCO. For the first edition, the variable is only provided for the years in which it was in use. For the second edition ASCO variable, it is provided for the pre-2009 years. For the pre-2003 period, the ASCO first edition codes are mapped to ASCO second edition codes using a concordance provided by the Australian Bureau of Statistics (ABS). The third variable is an ANZSCO 2006 variable which covers more recent years of data (1999 onwards). This variable is derived from occupation text descriptions in the tax forms using an auto-encoder provided by the ABS.

The fourth variable is a derived occupation variable which attempts to provide a consistent set of ANZSCO 2006 codes for all years, 1991 onwards. It uses ANZSCO 2006 codes from a combination of occupation codes and occupation text descriptions for the post-2008 period. ANZSCO 2006 for the pre-2009 years is created by mapping first and second edition ASCO codes to ANZSCO. There are three issues which arise in constructing this mapping. The first is that for some ASCO codes, there are no matching ANZSCO codes. For these observations (all pre-2009), the ASCO variable is non-missing but the ANZSCO variable is missing.[4] A second

[3] ANZSCO 2006 is also referred to as ANZSCO version 1.1. Version 1.3 is the most current version, released in November 2019.

[4] About 30 per cent of the ASCO occupation data is missing in the 2001 data; this goes up to 37.6 per cent for ANZSCO which include the cases where occupation codes could not be matched from ASCO to ANZSCO.

issue is that some ASCO codes can only be "partially" mapped to an ANZSCO code, that is there is not a one-to-one match. The ALife team used these partial mappings when generating the occupation variables. A third issue, and probably less important than the first two, is that the ABS did not produce a mapping from first edition ASCO codes to ANZSCO codes. Therefore, the ALife team, for the earliest years of the data (1991 – 2002), first mapped first edition ASCO codes to second edition ASCO codes and then mapped these to ANZSCO 2006 codes. The mapping thus will have likely created more errors in occupation in the 1991 – 2002 period than it does in the 2003 – 2008 period. It is nearly impossible to quantify the extent of this problem.

Table 1: Occupation variables available in the ALife data

Variable name	Years	Details
c_asco_fe	1991-2002	First 2 digits of first edition ASCO codes in the tax forms
c_asco_se	1991-2008	• Matching through concordance from first edition to second edition ASCO (1991-2002) • First 2 digits of second edition ASCO codes in the tax forms (2003-2008)
c_anzsco	1999-2017	First 2 digits of ANZSCO 2006 codes derived from occupation text descriptions in the tax forms
c_occupation	1994-2017	• Where possible, ANZSCO codes derived from text descriptions in tax forms (1991 – 2008) • Otherwise - First edition ASCO converted to ANZSCO 2006 (via second edition ASCO) (1991 – 2002) - Second edition ASCO converted to ANZSCO 2006 (2003-2008) • ANZSCO 2006 codes derived from a combination of occupation codes and text descriptions (2009-2017)

See text for more details

The fourth of these variables (`c_occupation`) is an attempt to provide an occupation variable with the least number of missing observations possible across the entire sample period. However, one consequence of the successive mappings of ASCO codes to ANZSCO codes is that the rates of occupation change in this fourth variable from 2002 to 2003 (when there was a change from first to second edition ASCO codes) and from 2008 to 2009 (when there was a change from ASCO codes to ANZSCO codes) might be higher than the actual rates of change. We discuss this in more detail below.

Table 1 presents a summary of the four occupation variables which are provided as part of the ALife data. In our analysis below, we consider the variable "c_anzsco" because it is consistently coded across the entire 17 years of ALife data that we use. We also consider the "c_occupation" variable as it has fewer missing values.

Comparing Occupation in HILDA and ALife

Table 2 presents the distribution of occupation for 2017, which is the most recent year available in both datasets. In all of our analysis we limit our sample to workers between the ages of 25-59, inclusive, as it reduces problems caused by those who are still pursuing an education, or transitioning from education to the labour market and those who are transitioning to retirement. This is particularly important when it comes to considering income from wages and salary since we only observe annual amounts in the tax data.

In both datasets, there are a substantial number of observations with missing information on occupation. In all of our results below, we focus on those for whom occupation data is non-missing. We do not address the question of the determinants of missingness in the data. ALife contains only very limited demographic information with which to address this question and this is not the focus of our paper. As indicated above, we focus on "c_anzsco" and "c_occupation" because it provides an occupation code with the least amount of missing data. Our a priori belief is that "c_occupation" will be more affected by the changes from ASCO version one to ASCO version two and ANZSCO.

The ALife sample is about 100 times larger than the HILDA sample.[5] In 2017, there are more managers and professionals in the HILDA sample and fewer technicians and trade workers and clerical and administrative workers.

"c_anzsco" is frequently missing in the ALife data. In the "c_occupation" variable, there is much less missingness. Only about 16.4 per cent of observations are missing for "c_occupation", as opposed to double that (36.6 per cent) for "c_anzsco".

Table 2: Distribution of Occupation in 2017

Occupation	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Managers	15.6%	13.3%	14.9%
Professionals	28.1%	27.4%	25.5%
Technicians and Trade Workers	12.8%	13.2%	12.4%
Community and Personal Service Work	11.5%	11.5%	10.6%
Clerical and Administrative Workers	12.9%	15.5%	14.6%
Sales Workers	5.1%	4.9%	6.7%
Machinery Operators and Drivers	6.4%	6.5%	6.1%
Labourers	7.6%	7.6%	9.3%
TOTAL (non missing)	8,063	638,878	842,457
Missing	20.0%	36.6%	16.4%
Total sample including missing	10,038	1,008,231	1,008,231

HILDA wave 17 (2017); ALife2017 (2016-2017 financial year)
Percentages in first 8 rows reflect the fraction of observations with non-missing occupation data
HILDA data weighted using cross-sectional, responding person population weights

In 2017, the ranking of most common to least common occupation in the dataset is quite similar. "c_anzsco" in ALife differs from the "c_occupation" variable and the HILDA data in the ranking for the second and third most common occupations (the clerical and administrative workers and managers is reversed). "c_occupation" in ALife shows more labourers and fewer managers than the other two data sources.

[5]In 2017 HILDA there are 8,166 individuals aged 25-59 with non-missing occupation data. When we apply the sample weights, 103 individuals with zero weight are dropped from the analysis, producing the sample size of 8,063 reported above. Throughout, we use weights and sample sizes reflect individual observations with non-zero weights.

Table 3 presents the same information for 2001, the first year for which we have data available in both datasets. Both datasets reflect occupation changes over time with managers having grown and the share of labourers having declined substantially. We also observe a large increase in community and personal service workers in both datasets. Missingness is more of a problem in 2001 than in 2017, particularly in the ALife data.

Table 3: Distribution of Occupation in 2001

Occupation	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Managers	14.5%	9.2%	9.7%
Professionals	24.8%	24.6%	25.2%
Technicians and Trade Workers	13.9%	14.5%	14.6%
Community and Personal Service Work	7.6%	8.1%	8.0%
Clerical and Administrative Workers	16.1%	19.3%	18.2%
Sales Workers	6.3%	6.9%	7.0%
Machinery Operators and Drivers	6.8%	5.9%	5.5%
Labourers	9.9%	11.5%	11.9%
TOTAL (non missing)	6,656	443,059	490,391
Missing	25.0%	43.6%	37.6%
Total sample including missing	8,891	785,953	785,953

HILDA wave 1 (2001); ALife2017 (2000-2001 financial year)
Percentages in first 8 rows reflect the fraction of observations with non-missing occupation data
HILDA data weighted using cross-sectional, responding person population weights

In our analysis below, we examine the occupational distribution and occupational change in conjunction with gross annual income and gross annual wages and salary. In both datasets there are some observations where people report zero wages and salary. In the ALife data, this is relatively rare. In 2001, 992 individuals of the 443,059 individuals in Table 2 (0.2 per cent) report zero wages and salary. In 2017, 0.9 per cent of individuals (5,905 of 638,878) report zero salary and wages. In HILDA, the number of people who report zero wages and salary is considerably larger. 757 individuals (11.4 per cent) in wave 1 and 681 individuals (8.5 per cent) in wave 17 report zero wages.

Tables 4 and 5 below report the occupational distribution in 2017 and 2001, respectively, for those with non-zero salary and wages. For the most part, the occupation distribution is unaffected. There are almost no changes for the ALife data and there are only minor changes in the HILDA data, most notably fewer managers and more professionals.

Table 4: Distribution of Occupation in 2017
(restricted to workers with non-zero annual salary and wages)

Occupation	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Managers	15.4%	13.2%	14.6%
Professionals	28.8%	27.5%	25.5%
Technicians and Trade Workers	12.0%	13.2%	12.4%
Community and Personal Service Work	11.6%	11.5%	10.6%
Clerical and Administrative Workers	13.4%	15.6%	14.6%
Sales Workers	5.2%	4.9%	6.7%
Machinery Operators and Drivers	6.6%	6.5%	6.1%
Labourers	7.0%	7.6%	9.4%
TOTAL (non missing)	7,382	833,202	833,202

HILDA wave 17 (2017); ALife2017 (2016-2107 financial year)
Percentages in first 8 rows reflect the fraction of observations with non-missing occupation data
HILDA data weighted using cross-sectional, responding person population weights

Table 5: Distribution of Occupation in 2001
(restricted to workers with non-zero annual salary and wages)

Occupation	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Managers	13.0%	9.2%	9.6%
Professionals	26.1%	24.6%	25.2%
Technicians and Trade Workers	13.4%	14.5%	14.6%
Community and Personal Service Work	7.9%	8.1%	8.0%
Clerical and Administrative Workers	16.7%	19.4%	18.2%
Sales Workers	6.4%	6.9%	7.0%
Machinery Operators and Drivers	6.9%	5.9%	5.5%
Labourers	9.7%	11.5%	11.9%
TOTAL (non missing)	5,899	442,067	489,431

HILDA wave 1 (2001); ALife2017 (2000-2001 financial year)
Percentages in first 8 rows reflect the fraction of observations with non-missing occupation data
HILDA data weighted using cross-sectional, responding person population weights

In what follows, we will focus exclusively on those with positive wage and salary income. In the Appendix, we reproduce the main results (tables and figures) including those individuals with zero wages and salary. The main conclusions that emerge from our analysis are unchanged if we include or exclude this group.

Occupational change

A key variable of interest is whether an individual changed occupations from one year to the next. We define occupation change equal to one if a person reported an occupation in two consecutive years and the occupations in the two years are different. Individuals who move from having no occupation to an occupation, or vice versa, are not counted as having changed occupations. This may slightly undercount the number of occupational changes, but our main interest is in comparing the two datasets so using a common definition across the two is of primary importance.[6] This approach ensures that we are comparing actual occupation changes rather than movements in and out of the labour force which might differ across the two data sources. Both HILDA and ALife would allow us to go back in time for people with no occupation to find out what their previous occupation was. This approach would be relatively unreliable in the ALife data given the larger amount of missing occupation data in earlier waves—see Tables 2 through 5. Finally, note that in general one can change occupation but stay with the same employer.

[6] People who move from occupation X to not working/no occupation and then to occupation Z will not be counted as having changed occupation. This will understate occupational changes if those who experience spells of non-working are more likely to change occupations when they come back to work than those who work continuously.

As we suspected, the rate of occupation change in the HILDA dataset is much higher than that in the ALife dataset as seen in Tables 5 and 6. Table 5 presents occupational change in the final wave of the data and Table 6 presents occupation change pooling across all years of data. Consistent with the job change results from Deutscher (2019), Tables 5 and 6 reflect that occupation change is lower in 2017 than it was on average over the 2001-2017 period. In addition, it is interesting that there is such a disparity between how often men and women report an occupational change in HILDA. In ALife, there is a small difference when using "c_anzsco", but in that case women report more occupational change than men. For "c_occupation" there are no differences in reported occupational change by sex.

Table 5: Percent Occupation Change in 2017

Variable	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Total Change	19.5%	5.4%	7.1%
Female	17.8%	5.7%	7.1%
Male	21.1%	5.2%	7.1%
Sample size	6,701	554,649	749,620

HILDA wave 17 (2017); ALife2017 (2016-2107 financial year)
HILDA data weighted using cross-sectional, responding person population weights
Only individuals with positive wage and salary income

Table 6: Percent Occupation Change across all years

Variable	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Total Change	22.1%	7.4%	9.7%
Female	19.6%	7.5%	9.5%
Male	24.2%	7.3%	9.9%
Sample size	95,625	7,367,396	10,001,767

HILDA waves 1-17 (2001-2017); ALife2017 (2001-2107)
HILDA data weighted using cross-sectional, responding person population weights
Only individuals with positive wage and salary income

To assess whether the change in occupation measured in HILDA was reasonable, we also looked at the Labour Force Survey (LFS) of the Australian Bureau of Statistics and longitudinal data from the Australian Census. The LFS data are monthly and individuals are surveyed in 8 consecutive months. There are 1,691,138 observations of individuals who are employed in consecutive periods and have an occupation listed in the period 2007 to 2019. We find monthly occupation switching rates from 10 to 19 per cent over these 12 years. In the Census, 38.6 per cent of people change occupation over a five-year period from 2006-2011 and 33.4 per cent from 2011-2016. These are based on individuals who are employed at both waves with non-missing industry and occupation. We conclude that the HILDA data look plausible based on these two independent data sources.

Why is there so much less occupational change in ALife than in HILDA?

As mentioned above, occupation change is lower in ALife because there is little incentive or requirement for individuals to report occupational changes to the Australian Taxation Office. In particular, when using electronic filing, occupation will generally be pre-filled from the previous year's return so it takes an active decision by the taxpayer to amend this field for an occupation change to be recorded.

In the following tables we compare occupation change for the two ALife variables for four sets of taxpayers: individuals who file electronically through a registered tax agent in two consecutive waves (ELS); individuals who file their own return electronically in two consecutive waves (ETAX); individuals who file their own return by paper in two consecutive waves (PPR); and those who switch between any of these three filing mechanisms.[7]

[7] ELS combines electronic lodgement by a tax agent with paper tax reforms which are filed electronically by tax agents; ETAX combines the current MYTAX electronic lodgement for individuals with the discontinued ETAX online lodgement; PPR are paper returns self-prepared by the taxpayer combined with a very small amount (less than one percent) of tax returns which are lodged by telephone.

The low rates of occupation change observed in ALife are driven by the low rates of occupation change for those who file electronically in consecutive years. The much higher rates for the switchers might be evidence that occupational switches may be associated with changes in how individuals file tax returns. This could be associated, for example, with switching to a tax agent due to an increase in wage and salary or business income or some other change in personal circumstances.

Table 7: Percent Occupation Change by tax return lodgement status (2016 - 2017)

	c_anzsco				c_occupation			
	ELS	ETAX	PPR	Mixed	ELS	ETAX	PPR	Mixed
Total Change	4.7%	5.3%	5.0%	10.2%	6.4%	6.4%	6.1%	13.7%
Female	4.9%	5.6%	4.8%	10.6%	6.3%	6.6%	5.8%	13.6%
Male	4.5%	4.9%	5.2%	9.8%	6.4%	6.3%	6.5%	13.7%
Sample size	554,649				749,620			

ALife2017 (2016-2107 financial year)

Table 8: Percent Occupation Change by tax return lodgement status (all years)

	c_anzsco				c_occupation			
	ELS	ETAX	PPR	Mixed	ELS	ETAX	PPR	Mixed
Total Change	5.5%	10.5%	9.6%	15.9%	7.6%	13.0%	11.0%	19.4%
Female	5.6%	10.3%	9.7%	16.1%	7.4%	12.3%	10.8%	18.8%
Male	5.4%	10.8%	9.5%	15.7%	7.8%	13.8%	11.2%	19.9%
Sample size	7,367,396				10,001,767			

ALife2017 (2001-2107)

As we can see in Tables 7 and 8, changing lodgement status is correlated with higher reports of occupational change. Individuals (ETAX and PPR) are much more likely to report occupational change than tax agents (ELS). This would be consistent with tax agents not asking clients about occupation but simply filling in the information from the prior year. Surprisingly, paper returns produce slightly lower reports of occupational change than electronic filings for individual taxpayers. Perhaps being prompted by the dropdown menu at the beginning of the electronic filing process prompts people to report occupational changes whereas those filing paper returns might simply copy occupation from last year's return.

Demographic variables

We are interested in how the distribution and mobility of occupation vary across age. It is important to note that since we have only 18 waves of data in HILDA and 27 in ALife, the full span of a worker's time in the labour force is not covered by our available data. In addition, we are not able to see intergenerational trends in occupational mobility. In some of the graphs presented below we create 5 age groups each of which span 7 years. In 2017, the mean age of our sample is 40.6 in the ALife sample and 41 in the HILDA sample—see Table 9. This difference is not economically significant.

Table 9 presents other descriptive statistics comparing the HILDA sample to the ALife sample. For the "c_occupation" sample in ALife, the numbers are almost identical to the "c_anzsco" sample.[8] Marital/partnered status and the presence of dependents may have an influence on occupational mobility. We can see that in the final wave of data over 4% more of the sample are married in the ALife Dataset than in HILDA. Note that in the tax data, marital status is determined by a taxpayer providing required information about a spouse or de facto partner. We cannot distinguish between marriage and partnership. Officially, individuals are supposed to provide information about their partner's income if they co-reside with someone "on a genuine domestic basis in a relationship as a couple"[9] and have co-resided for more than six

[8] These and any other results which are discussed but not presented are available from the authors upon request.
[9] From the ATO website (https://www.ato.gov.au/Individuals/myTax/2021/In-detail/Personalise-return/?anchor=Didyouhaveaspouse): "ATO considers a spouse anyone you've lived with in a genuine domestic relationship at any point during the year. Your spouse includes another person (of any sex) who:

months. We do not know anything about the levels of compliance with this rule. The number of dependents is quite similar across the two datasets.

Table 9: Demographic Variables in 2017

Variable	HILDA	ALIFE
Percent Male	52.1%	51.0%
Percent Female	47.9%	49.0%
Mean Age	41.0	40.6
Percent Married	58.0%	n/a
Percent Partnered	73.6%	59.7%
Mean # of Dependents	0.95	0.92
Mean Gross Income	$84,716	$74,212
Mean Wage and Salary Income	$75,421	$66,843
Sample size	7,384	833,202

HILDA wave 17 (2017); ALife2017 (2016-2107 financial year)
HILDA data weighted using cross-sectional, responding person population weights
Only individuals with positive wage and salary income

For our analysis of the distribution of occupation and occupational mobility across income groups we use two different measures of income: financial year gross total income and financial year gross salary and wages. For both measures we compare the quintiles of the distribution. First, we create gross income quintiles. This measure of income includes workers' gross income from labour, investments, and private and public transfers. The mean gross income of the sample in the HILDA data is $84,716 while the mean gross income in our ALife sample is $74,087. In addition, we create quintiles based on yearly wage and salary income as it may be more closely linked to occupational choices and changes. In the HILDA sample the mean wage and salary income is $75,421 while it is $66,843 in the ALife sample. We are not sure why there

- you were in a relationship with that was registered under a prescribed state or territory law
- although not legally married to you, lived with you on a genuine domestic basis in a relationship as a couple.

If you've entered into a spousal relationship during the year, or separated from your spouse part way through, you'll need to tell us the relevant dates so we can accurately calculate these amounts."

is such a difference in both gross income and salary and wages between the two datasets. If we include the zero salary and wages (see Appendix Table A3), the difference narrows somewhat but we still see higher reports of income in the HILDA data. This could be driven by elements of income that are non-taxable and do not appear in the ALife data.

The sex composition is slightly different among the two datasets, with the HILDA data being slightly more male-dominated. In the raw data there are more women than men, but the weighting corrects for the fact that non-respondents in HILDA are more likely to be men—see Summerfield, et. al. (2019). We examine occupation separately by sex in what follows.

Table 10 presents information on the labour force status of HILDA respondents in 2017. The ALife data contain annual wage and salary information. One can infer that individuals worked at least some part of the year if they received wages or salary, but it is impossible to infer whether the work was full- or part-time. Similarly, for those who don't work, we cannot distinguish between those who are unemployed and those who are not in the labour force.

Table 10: Labour Force Status in 2017

	HILDA		ALIFE	
Labour Force Status	**Male**	**Female**	**Male**	**Female**
Full Time	89.6%	60.2%	N/A	N/A
Part Time	10.3%	39.7%	N/A	N/A
Unknown	0.13%	0.13%	N/A	N/A
Sample size	3,743	3,639	N/A	N/A

HILDA wave 17 (2017)
HILDA data weighted using cross-sectional, responding person population weights

In thinking about factors that may contribute to occupation choice and occupation change, it is also important to look at the geographical distribution in the two datasets. Geographical location may tie individuals to a specific occupation or heavily influence their occupational choices. The ALife and HILDA data are very similar once the HILDA weights are applied as can be seen in Table 11.

Table 11: Geographic Variables in 2017

Geographic Area	HILDA	ALIFE
NSW	31.5%	31.9%
VIC	27.8%	25.5%
QLD	19.3%	19.6%
SA	6.4%	6.6%
WA	10.0%	10.9%
TAS	2.0%	2.0%
NT	1.1%	1.1%
ACT	1.9%	2.0%
Other	0.0%	0.0%
Major City	72.0%	75.6%
Inner Regional	19.6%	15.8%
Outer Regional	7.3%	7.3%
Remote Area	1.1%	1.0%
Very Remote Area	0.1%	0.3%
Urban Area	84.4%	89.0%
Not Urban	14.5%	11.0%
Sample size	7,382	833,202

Most of the differences that we observe between the two datasets arise from the different nature of the two datasets. ALife shows much less occupational change than HILDA because taxpayers do not need to report occupational changes in order to have their tax liability assessed. HILDA, with its strong focus on labour market dynamics, includes a battery of questions to determine whether an individual has changed jobs or occupation. The distribution of occupation in ALife is slightly biased towards slower growing occupations again because ALife is not tracking occupational change very well.

Gross income is larger in HILDA because people might include income in their survey responses that doesn't need to be reported to the tax office. ALife also misses some people who have little or no wage and salary income and who fall below the tax-free threshold. ALife will also miss those who work in family businesses and are not paid a salary or some type of business income.

Average salary and wage income is larger in ALife than in HILDA when we include those with zero salary and wages (most of these are in the HILDA data). Average wage and salary income is larger in HILDA than in ALife when we drop the zeros. It could be that those who receive small, non-zero amounts of wage and salary are better captured in ALife than in HILDA. Some of the zero reports in HILDA may in fact be people with very small amounts of salary and wages who mistakenly report zero. An interesting topic for a future research paper similar to this one would be to examine income differences in HILDA and ALife across a range of socio-demographic dimensions.

IV. The distribution of occupation

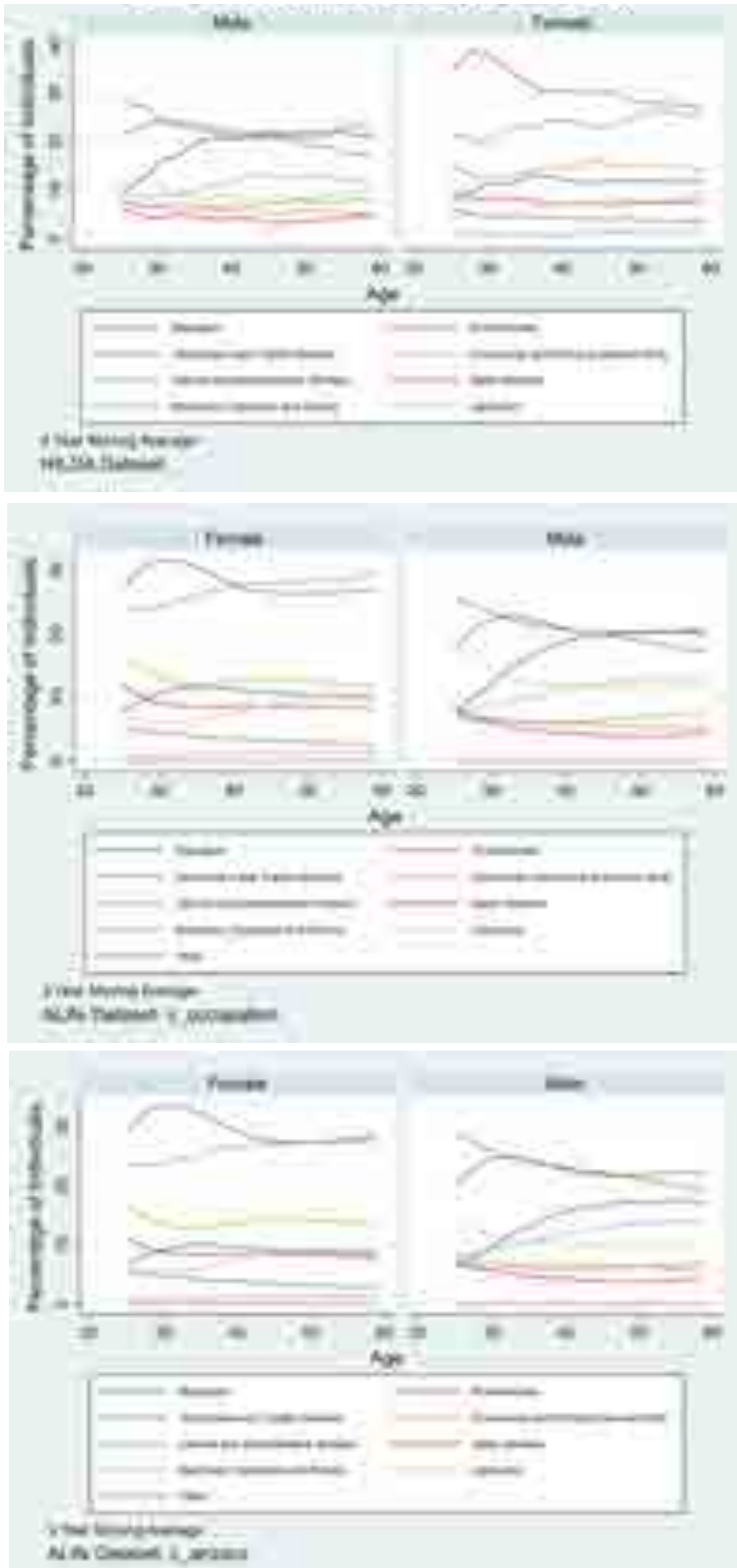
Next, we graphically examine the differences in the distribution of occupation and occupation change between the two datasets. In what follows, we create a series of graphs that show the percentage of individuals in the sample that were employed in one of the eight occupations, as well as a second series that shows the percentage of individuals that changed occupation. Separate analysis by sex allows us to examine whether trends in occupation are the same for men and women. We look at both the distribution of occupation and frequency of occupation change across age groups, gross income quintiles, and wage and salary income quintiles. In all our graphs, we use three-year moving averages. We find this to be helpful in reducing the amount of noise that is present, particularly for the HILDA data. The moving average makes comparison with ALife, which has much less volatility, easier. For the main analysis, we drop those who report zero salary and wages. We repeat much of the analysis in the Appendix where we include those observations with zero salary and wages. Our main conclusions are unaffected.

We also examine the frequency of occupation change by occupation type. To view more macro trends, we also look at the total amount of occupation change over time and age. Lastly, to connect our analysis of occupation to questions of income we look at the mean of both gross income and wage and salary income across type of occupation over both time and age.

Distribution of Occupation by Age

Looking at the distribution of occupation across age and sex in Figure 1, it is remarkable how similar they are across the two datasets. The only noticeable difference is that for the ALife dataset, clerical and administrative workers are the most prevalent for women over 40, and in HILDA, professionals are the most prevalent across all female workers. In addition, for female workers the distribution of occupation becomes more compressed in the HILDA dataset over age than it does in ALife.

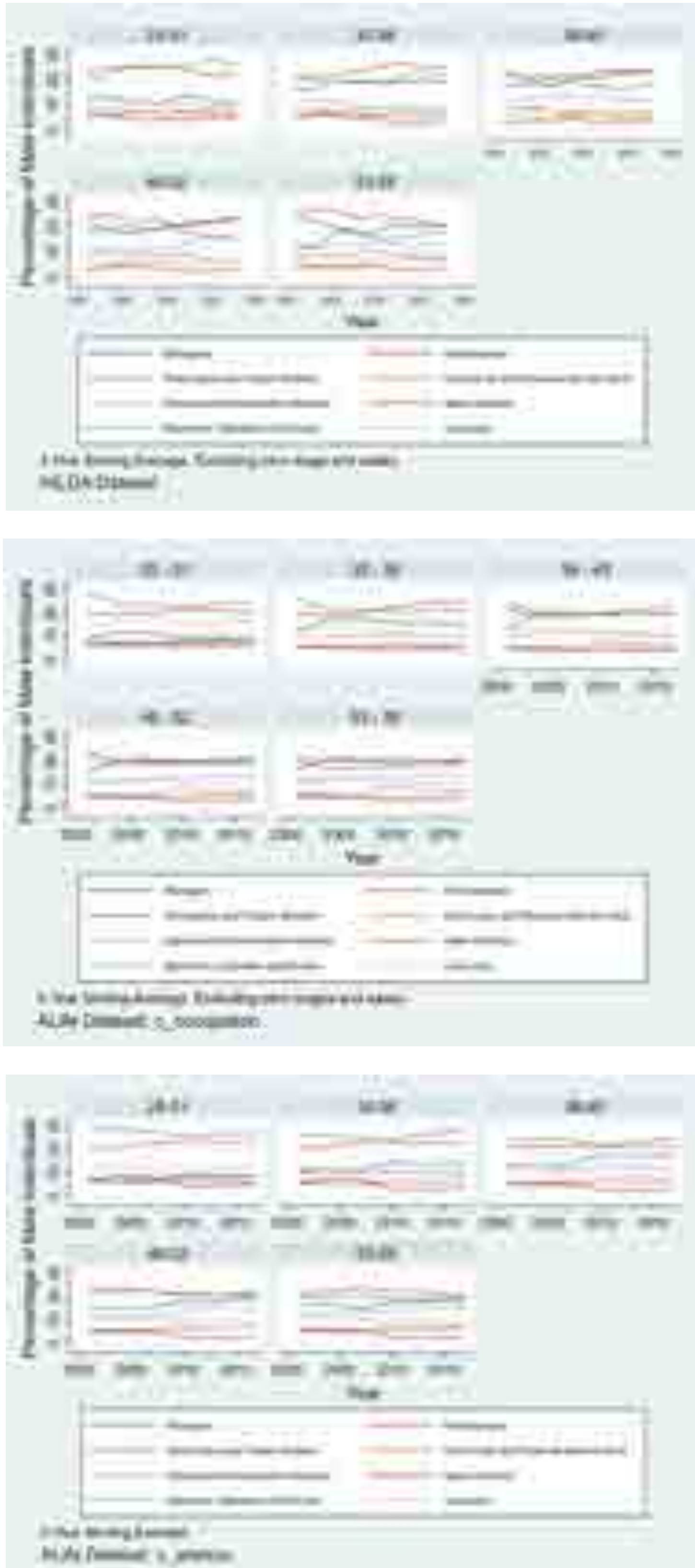
Figure 1: Distribution of occupation by age and sex



In Figure 2, we can see that, for men, the trends in occupation in the two datasets are relatively similar across time and age groups. The difference that is most apparent is that the distribution is much smoother over time in the ALife dataset which makes sense given the fact that the sample is much larger and the rate of reported change is much smaller. In addition, we can see that there is a difference in the first five years of the distribution between the the ALife and HILDA datasets. In ALife there seems to be a large decrease in the share of technicians and professionals and an increase in the share of managers before the year 2005 that is not apparent in the HILDA sample. This seems consistent with wider trends away from manual work towards brain work.[10] In both of these figures, we can see that as the Australian workforce gets older, the share of technicians and labourers decreases, while the share of professionals, managers, and machinery operators and drivers increases. The distributions are relatively constant over time.

[10] We note that the managers group includes the sub-category “Farmers and farm managers”, but this is only a small proportion of those who are managers.

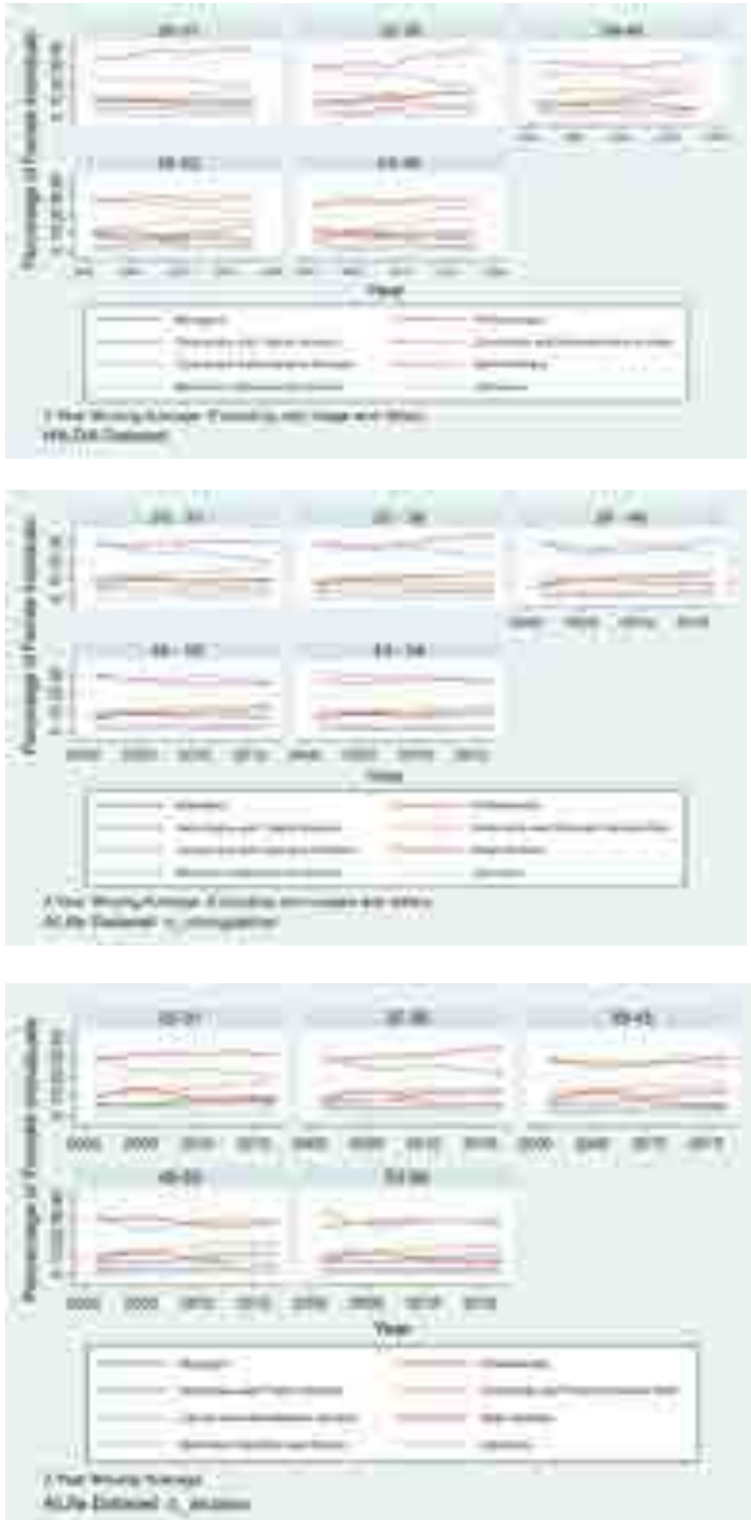
Figure 2: Distribution of occupation by age group: males



For females, we see larger differences between the two datasets in occupation trends over time by age group—see Figure 3. In general, professionals and clerical workers dominate the distribution across all age groups for both datasets. In our HILDA sample, the share of professionals is increasing over time, while the share of clerical workers has been decreasing. In addition, we can see that as age increases, the gap between the share of professional workers and clerical workers closes. In addition, community workers and managers increases both over time and across age groups.

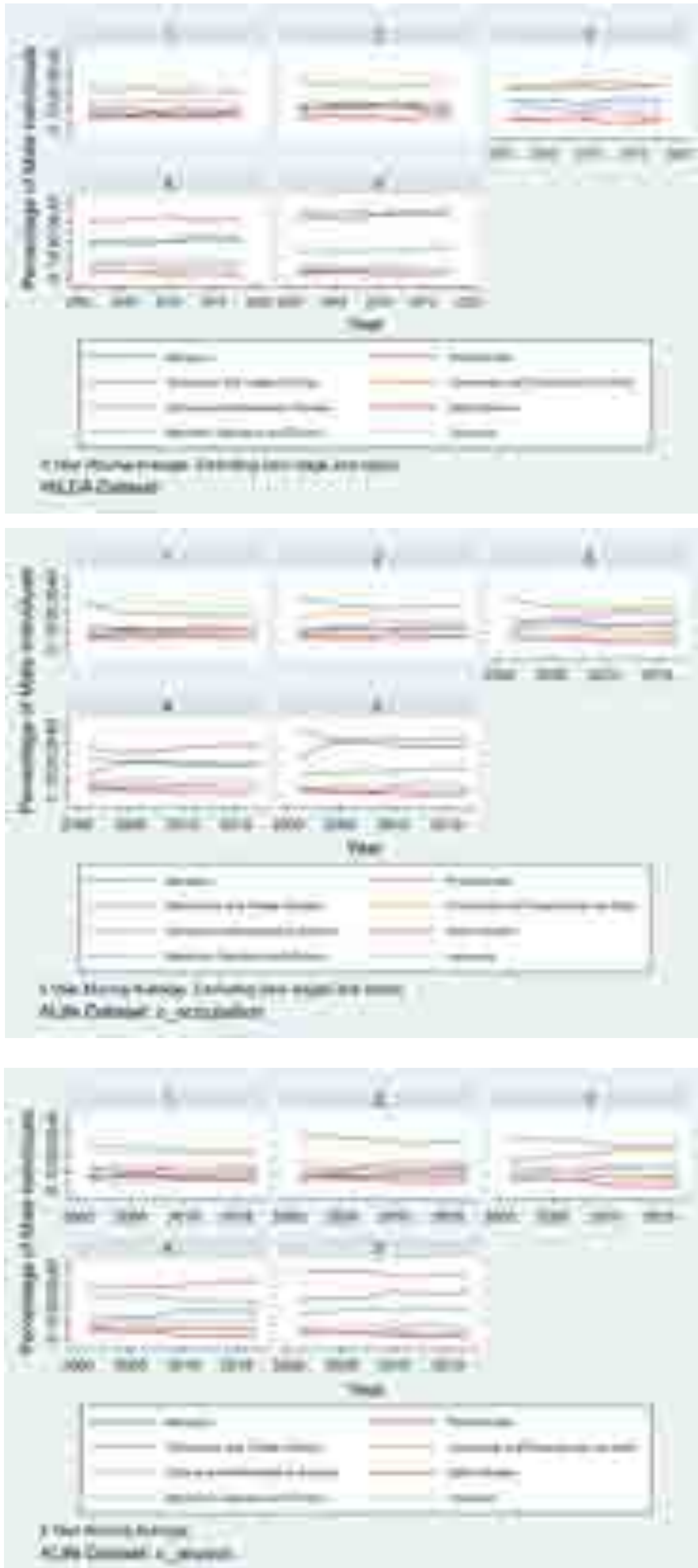
The ALife sample shows a similar trend only for the ages of 25-38. In the ages of 39-59, it appears that clerical workers and professionals have had a consistent share of around 30 per cent of women's occupations while the other occupations have all been lower than 15 per cent. In fact, clerical workers were more prevalent in the 39-59 age groups than professionals. It is important to note that the graph based upon the HILDA data has a vertical axis that is 10 percentage points higher than that of the ALife data—demonstrating that the spread in occupation is much greater as is the dominance of the presence of professional across all age groups.

Figure 3: Distribution of occupation by age group: females



Distribution of Occupation by Income Quintile

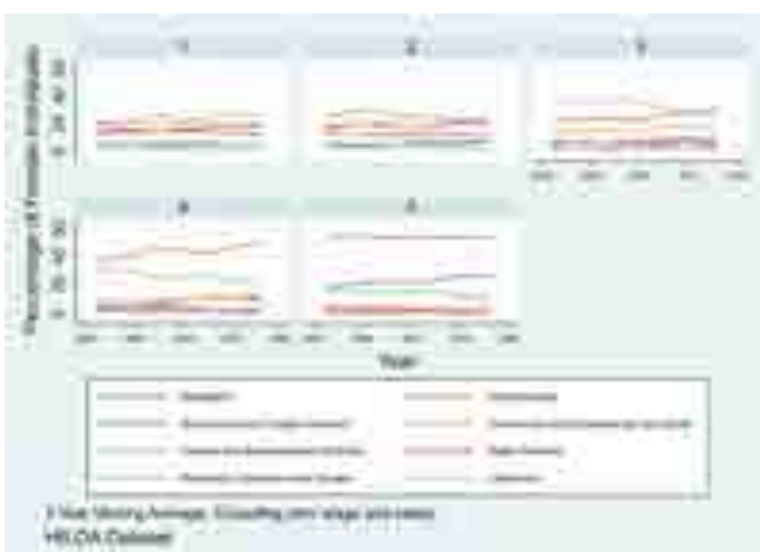
Figure 4: Distribution of occupation by gross income quintile: males

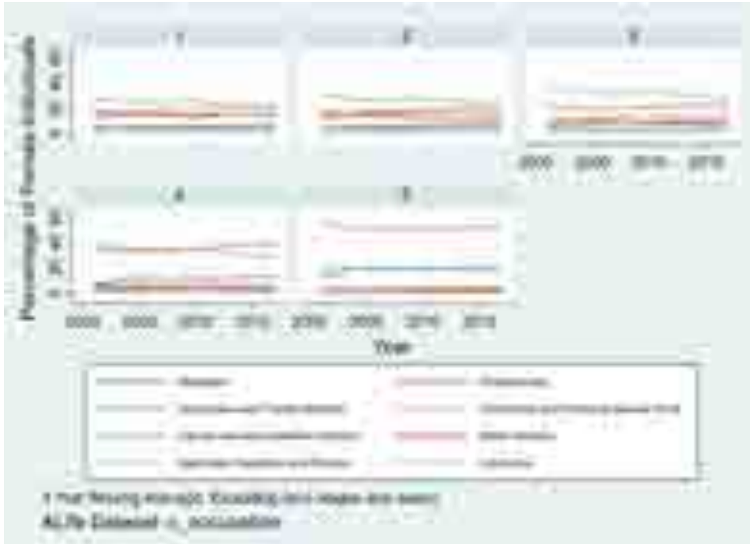


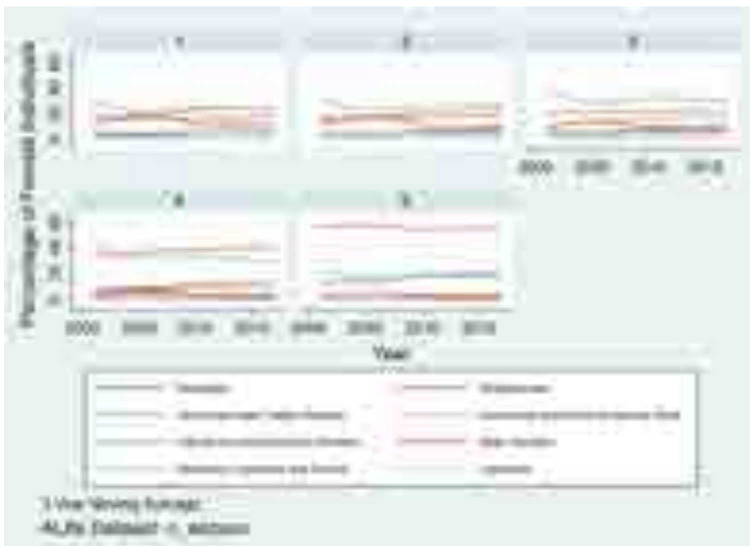
When looking at the distribution of men's occupations across gross income quintiles in Figure 4, we find greater differences between the two datasets. Most notably in the lowest income quintile, the distribution of occupation is much more dominated by labourers in the ALife dataset, while technician and trade workers are the most prominent in HILDA. In the next four income quintiles, the distributions are similar across the two datasets. While they are constant over time, it is apparent that as income rises, the share of technicians and trade workers and labourers fall, while the share of managers and professionals rises.

Looking at Figure 5, in the bottom income quintile for women, clerical workers, community and personal service workers, and professionals are the most prevalent across the two samples. We can see that as income rises there is a widening in the distribution such that in the highest income quintile, professionals compose almost 60 per cent of occupation in both samples. Also, as income rises, the share of professionals and managers increases. At the same time, clerical workers increase with income till the highest income quintile, but also decrease over time in all income groups. Community and personal service workers are very prevalent in the bottom two income quintiles.

Figure 5: Distribution of occupation by gross income quintile: females

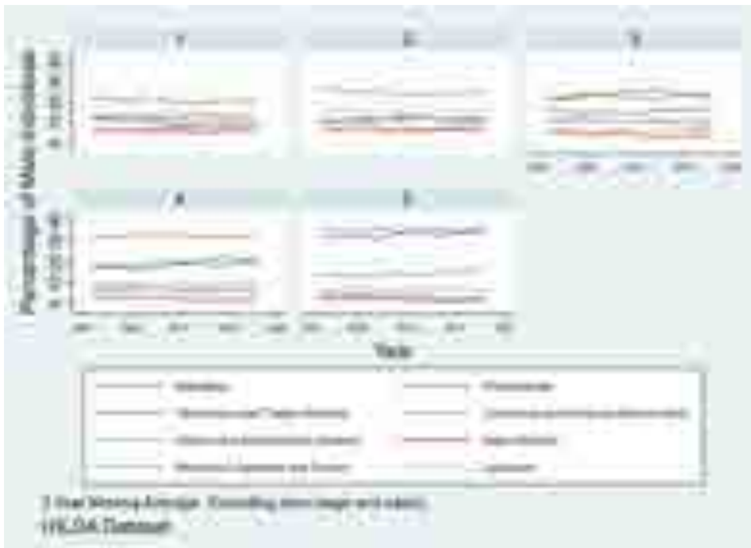


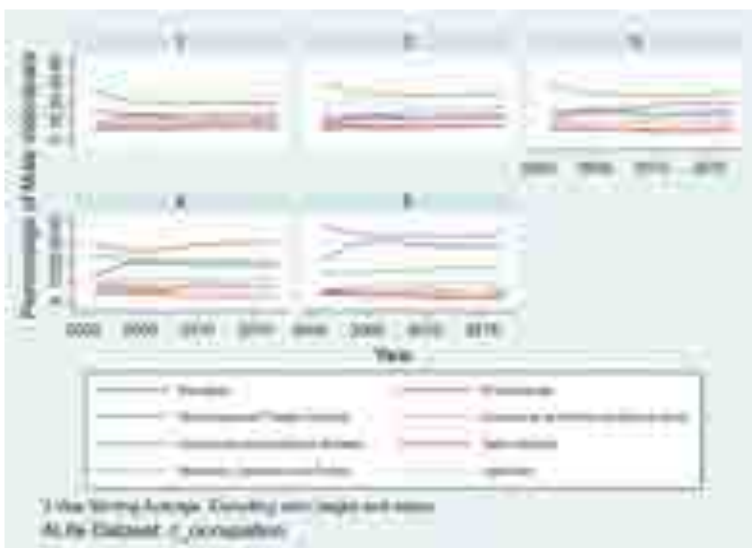


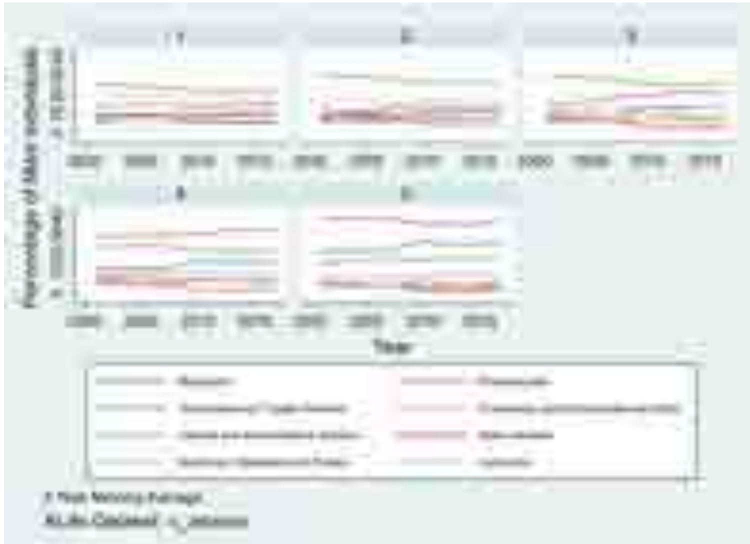


In looking at the distribution of occupation across wage and salary income, as opposed to gross income, we can again see some differences and some common patterns. For men, in Figure 6, both samples show that as income rises, the share of labourers and technicians falls, while the share of professionals and managers rises. Yet there are several differences in the two distributions. In the second quintile, while technicians and trade workers are the most prevalent workers in both datasets, machinery workers are the second most in HILDA and labourers are second most in ALife. In the third quintile, professionals are far less prominent in the ALife sample than in the HILDA sample. Lastly, in the top two quintiles managers are less prominent in ALife than in HILDA. Over time, the distribution of occupation seems relatively flat in both datasets.

Figure 6: Distribution of occupation by wage and salary income quintile: male

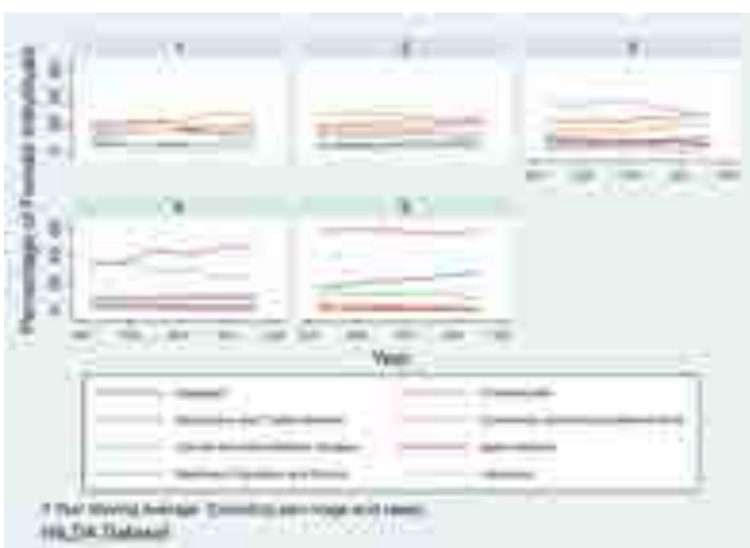


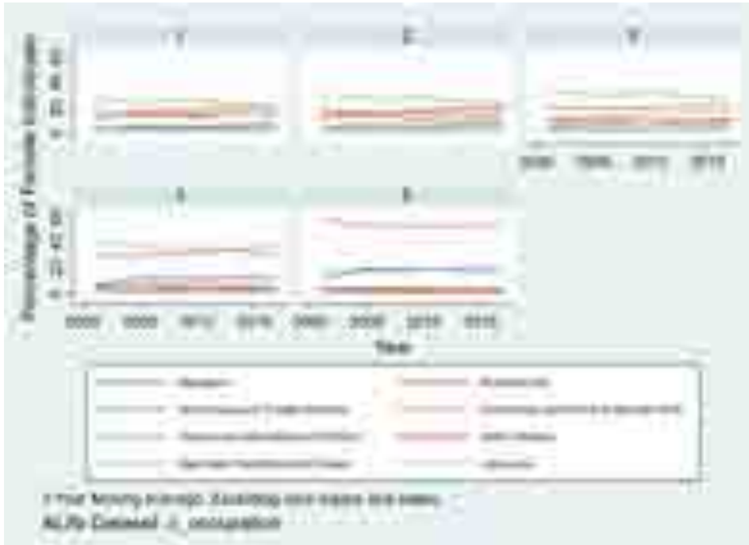


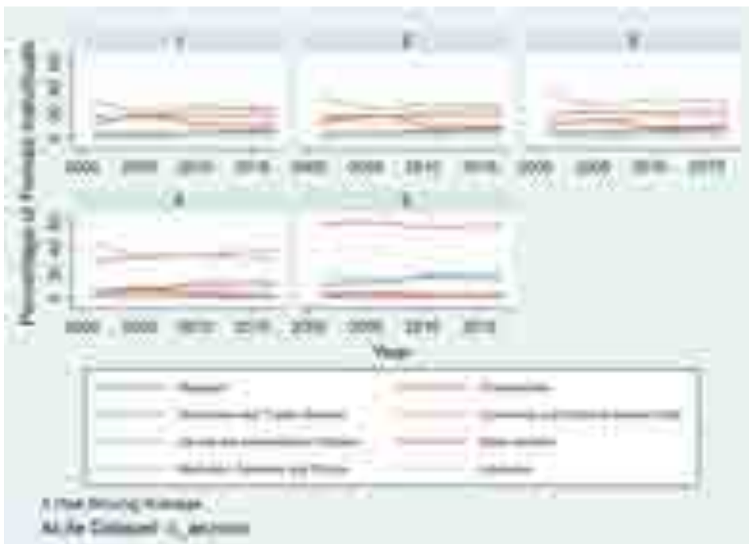


For women, in Figure 7, we again see that as income rises the distribution of occupation becomes more concentrated in certain occupations. The share of clerical workers rises with income till the fourth quintile, then drops back down in both samples. Meanwhile we again have the share of professionals rising from about 15 per cent in the bottom quintile to about 60 per cent in the highest quintile. Community and personal service workers are highly prevalent in the first three income quintiles for women in both samples.

Figure 7: Distribution of occupation by wage and salary income quintile: female







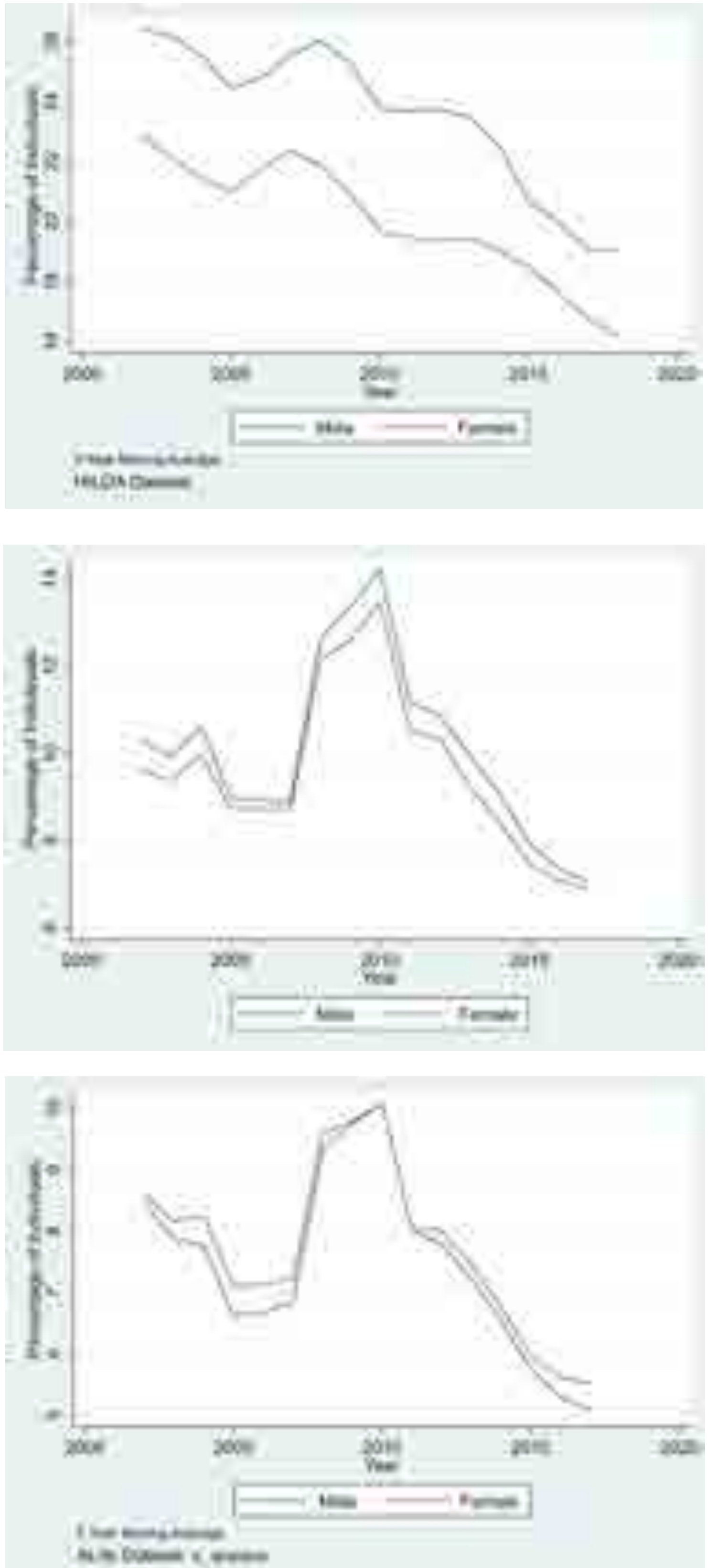
There are also some differences between HILDA and ALife. In the first quintile, community service workers are the most prevalent in HILDA, while clerical workers are the most prevalent in ALife. The second and third income quintiles are very similarly distributed for the two samples. In the fourth income quintile, professionals are not as prevalent in ALife as in the HILDA data. Lastly, in the top income quintile we can see that in ALife there is a smaller disparity between the proportions of clerical workers and managers.

V. Occupation Change

We first look at occupation change over time in the two datasets with the data split by male/female. There are three striking things from Figure 8. First, men change occupations more frequently than women in the HILDA data but there is little difference in the ALife data. Second, there is a strong downward trend over time in occupation change. Third, there is a large spike in occupation change during the Global Financial Crisis (GFC) that is visible in the ALife data but not in the HILDA data.

If we look at occupation change by age for men and women, we also see a pattern that occupation change decreases as people age. Again, we see that there appears to be a difference between men and women in HILDA, but not in ALife. In fact in ALife, there are some ages where women appear to change occupation more frequently than men. This could be driven by a reporting effect. The observed pattern is consistent with a story where women are more likely to update their occupation details with the Australian Taxation Office but both men and women are equally likely to report occupation changes on the HILDA survey. There are other possible explanations but it is difficult to distinguish between competing stories without additional data or without some type of matched data between the two data sources.

Figure 8: Occupation change over time by sex

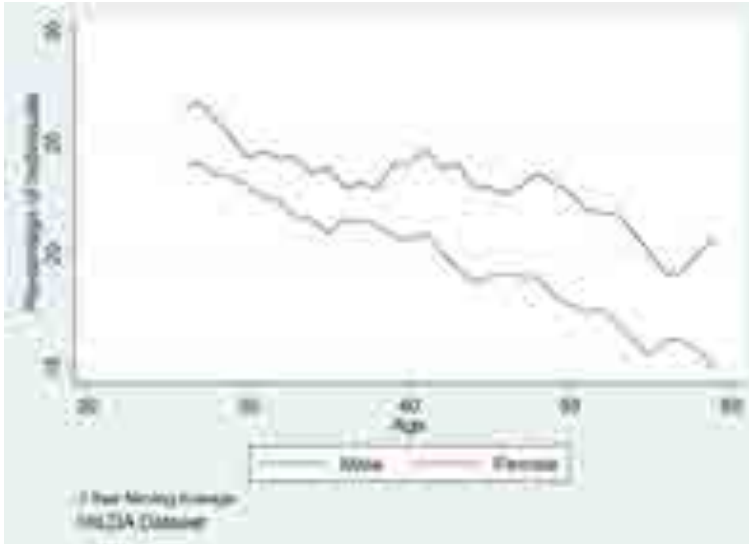


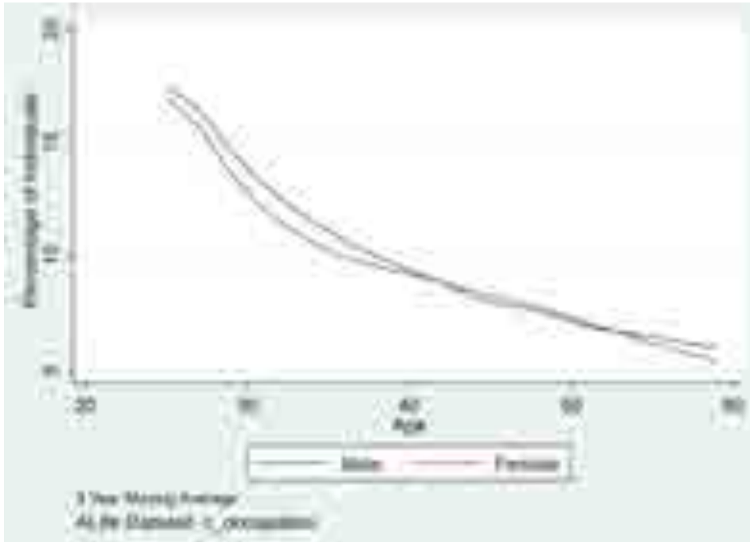
Appendix Figures A5 – A8 show this relationship for the ALife data, but with individuals split by lodgement type. The big spike in occupational change during the GFC is noticeable in all lodgement types but is particularly pronounced in those returns that are filed by tax agents. Tax agents may have been aware that the occupation coding reported to the ATO had changed and they may have induced changes in occupation either by asking clients about occupational changes more than they normally would or by revising people's occupations in a way that inadvertently provoked a change in the coded occupation.[11]

Figure 9 looks at the rate of occupational change by age. We can see that there is much more volatility in the HILDA dataset even when using the three-year moving average. In addition, we again see that the rate of occupation change is much higher in the HILDA sample. In HILDA we also see that men change occupation more frequently than women. This pattern is not reproduced in the ALife data where the patterns for men and women across age are very similar.

[11] Another possibility is that the ATO updates the electronic lodgement service periodically and for years where the occupation coding changed, there may be no previous occupation information available to taxpayers and tax agents which is pre-filled in the electronic tax return.

Figure 9: Occupation Change over age by sex





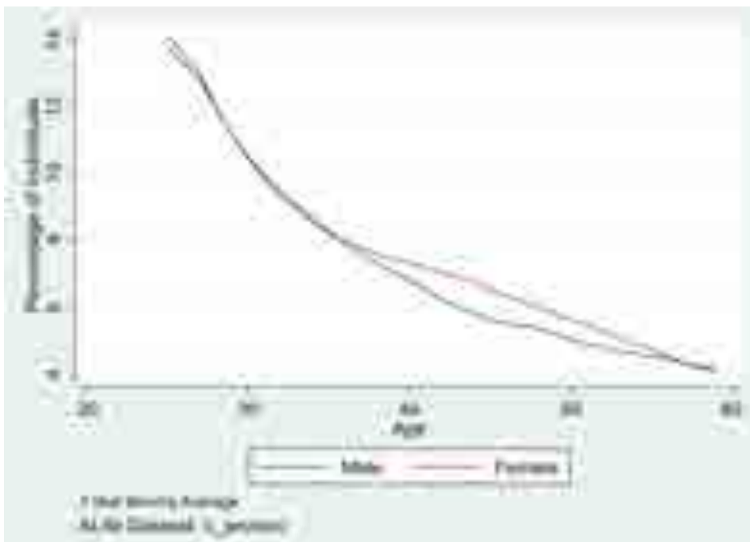
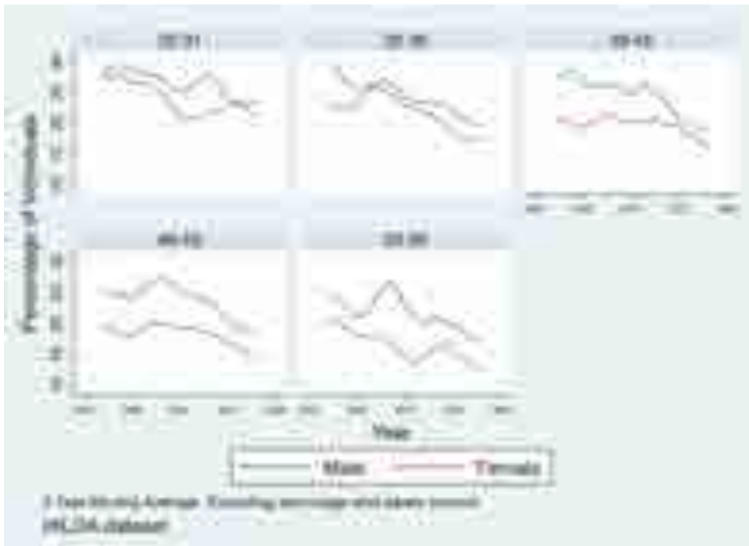
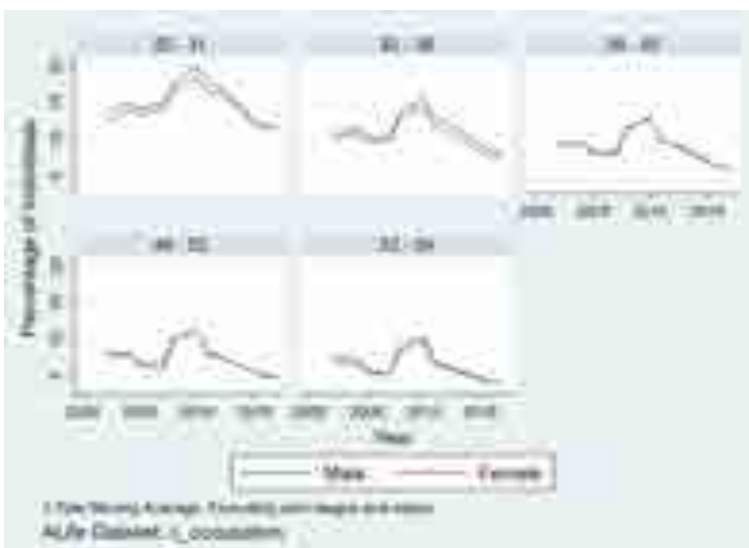
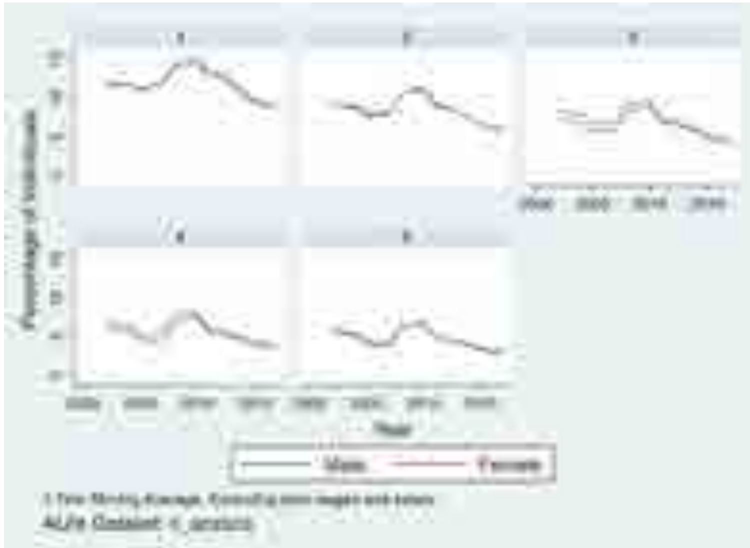


Figure 10 presents occupational change, by age group, split by male/female. Beyond the already observed differences in the average rate of occupational change in the two datasets, this figure reveals stark differences in the evolution of occupation change across time. As we saw in Figure 8, the ALife data show a sharp jump up in occupational change during the Global Financial Crisis (GFC), quite different from the HILDA data. This large spike in occupation change in 2009-2010 in the ALife data, coinciding with the GFC, is visible for all age groups. HILDA appears to show something similar for 46-59 year old males, but not for females. For the 25-31 and 39-45 year age groups, there seems to be a sharp increase in occupational change several years after the GFC, again for men. For women, in HILDA, we observe a sharp drop in occupational mobility in the 25-31 year old age group around the time of the GFC.

Figure 10: Occupation Change by age group

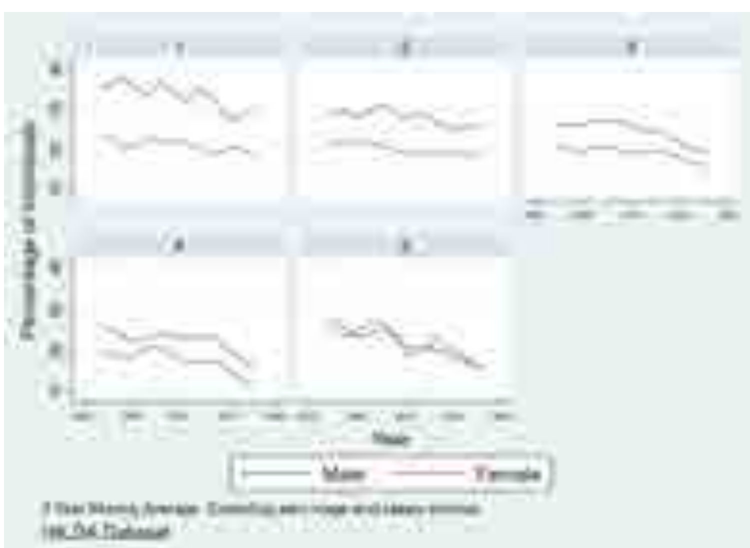


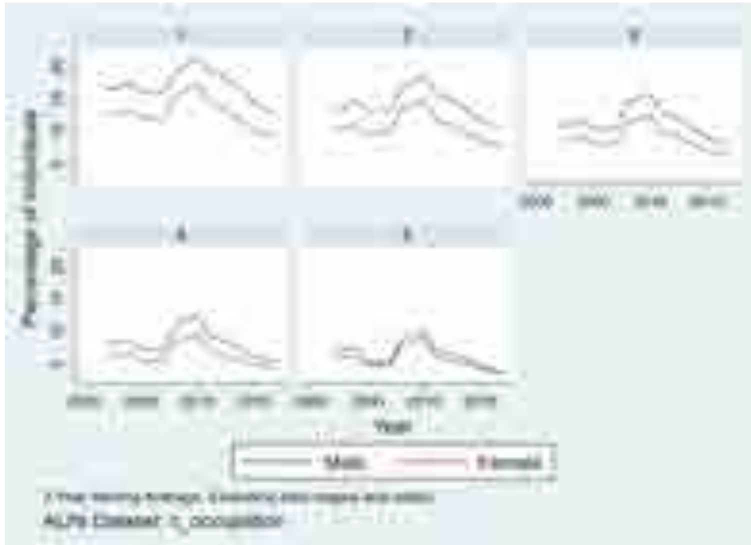


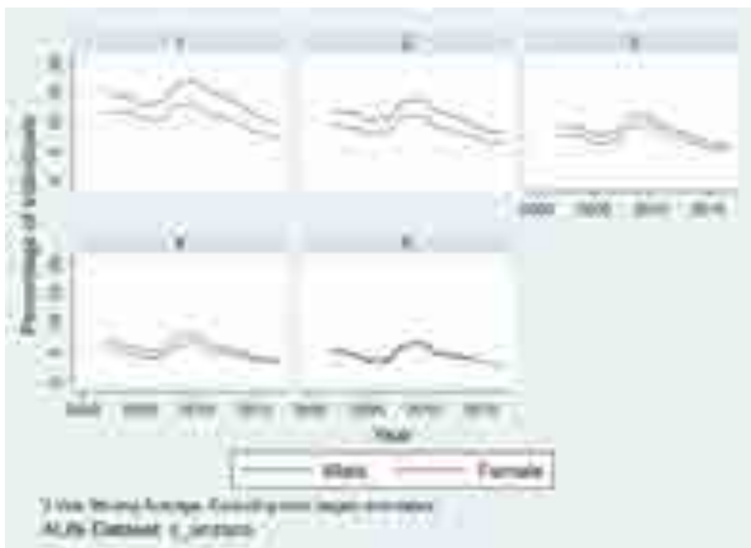


In both datasets, the overall trend is of decreasing occupational mobility over time although the decrease appears sharper in the HILDA data than in the ALife data. Figures 11 and 12 show the trends over time for occupational change for men and women, but now split by income quintiles. Figure 11 uses gross income to form the quintiles whereas Figure 12 uses wage and salary income.

Figure 11: Occupation change by gross income quintile

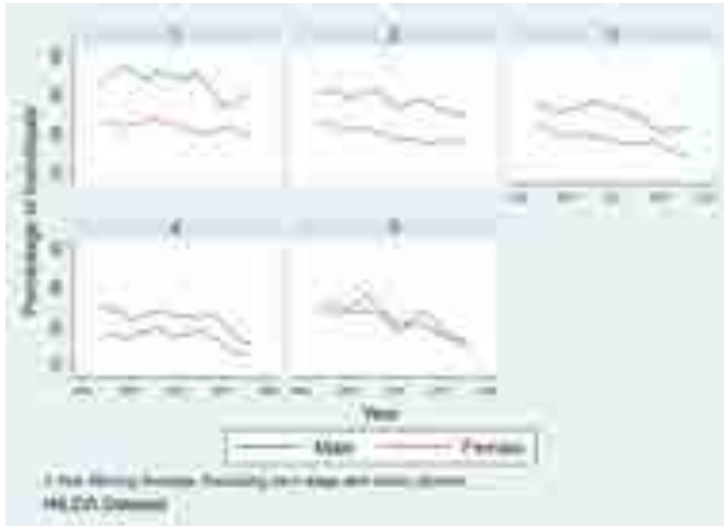


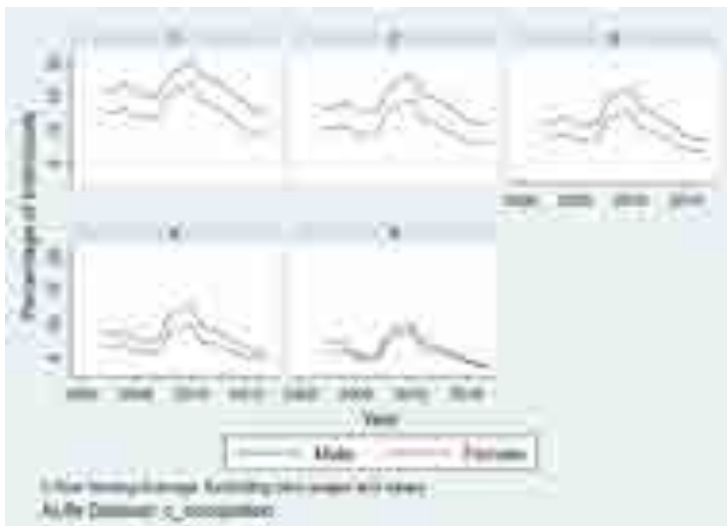




Both figures show the same trends and in both we see less mobility at higher income levels, consistent with the research of Sicherman (1990), reported above. Interestingly, in the HILDA data, there are a couple of years where women change occupation more than men with big spikes of occupation change in 2007 and 2013. We have no idea what drives this.

Figure 12: Occupation change by wage and salary income quintile





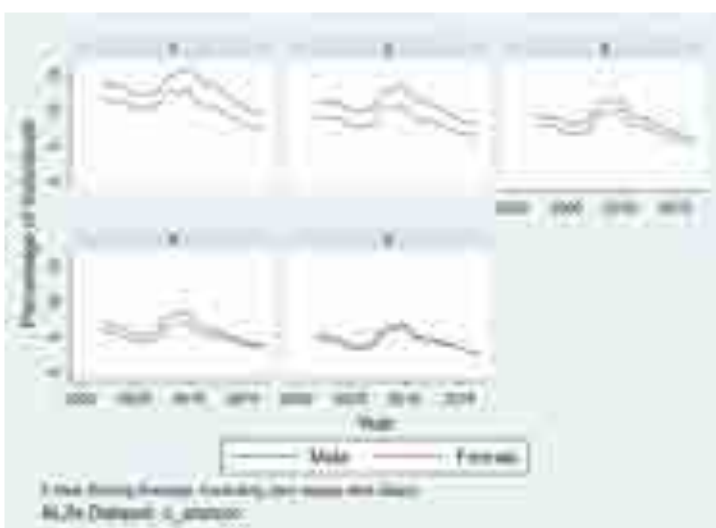
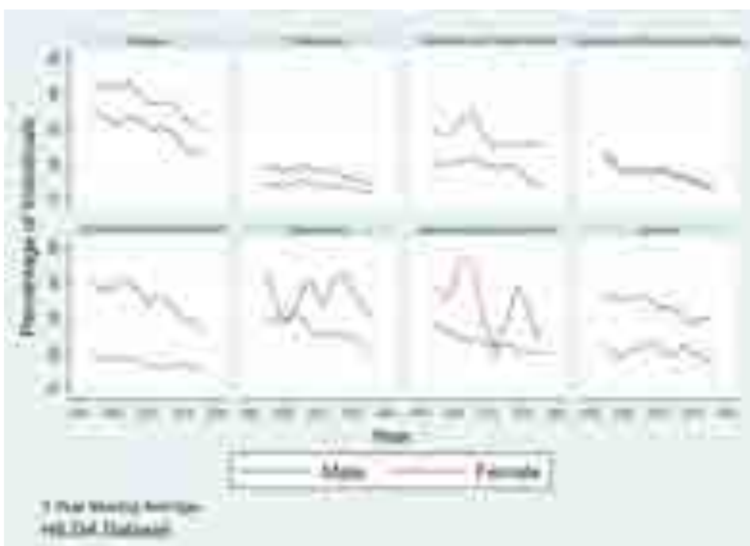
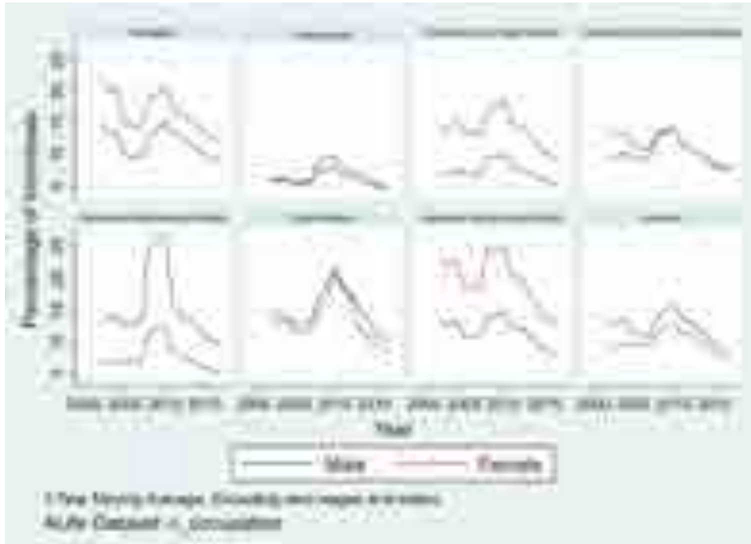


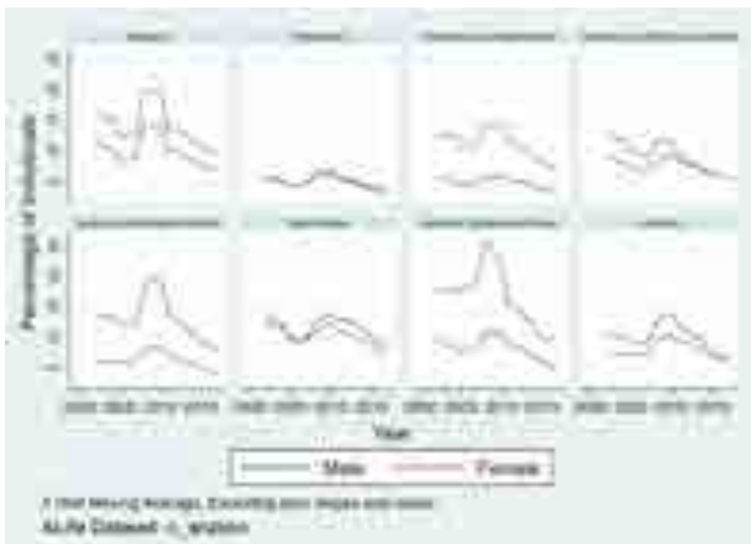
Figure 13 looks at occupation change by occupation type. The trends over time of decreasing occupational change seem present in all occupations as do the spikes in occupational change around the GFC seen in the ALife data. Some occupations (managers, technicians and trade workers and machinery operators) have higher occupational change than some others (professionals and labourers). These are consistent across the two datasets.

There are some interesting gender patterns. Looking at occupation change overall, we generally found that men changed occupation more than women. But when we look by occupation, we can see that women managers, technicians and trade workers and machinery operators and drivers are much more likely to change occupation than men. These patterns are consistent across both datasets.

Figure 13: Occupation change by occupation type



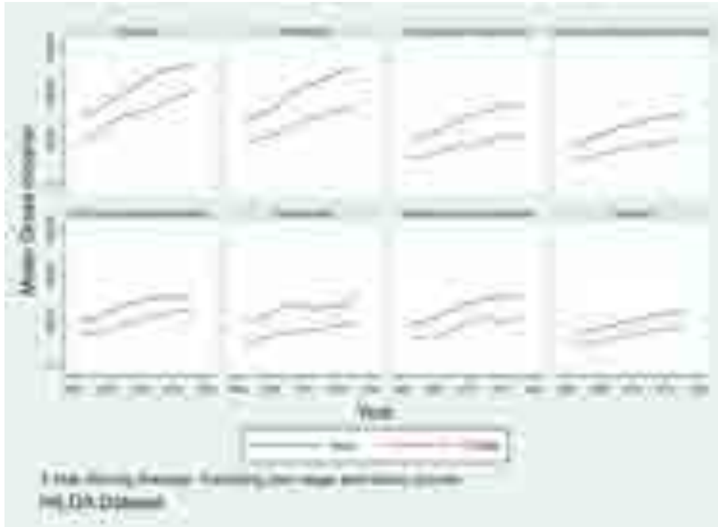


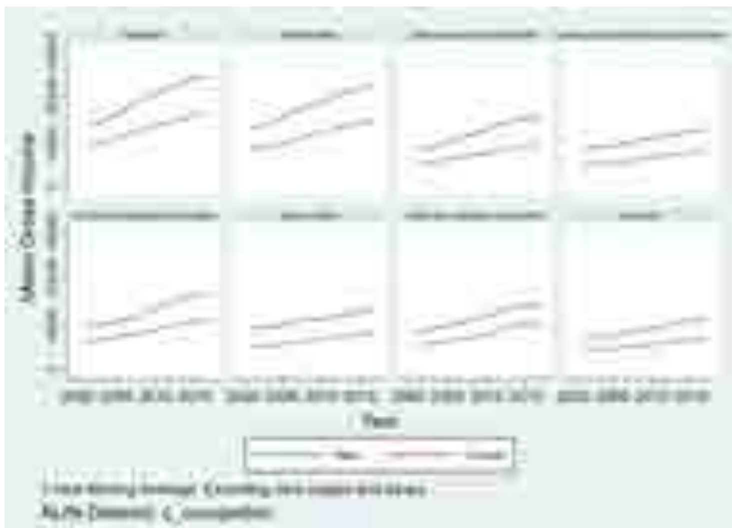


Income and wages by occupation type

Figure 14 shows that the overall patterns of gross income by occupation type from HILDA and Alife are very similar, and that men earn more regardless of occupation type. This obviously doesn’t account for hours worked and we know from Table 10 that women are much more likely to be part-time.

Figure 14: Mean gross income by occupation type





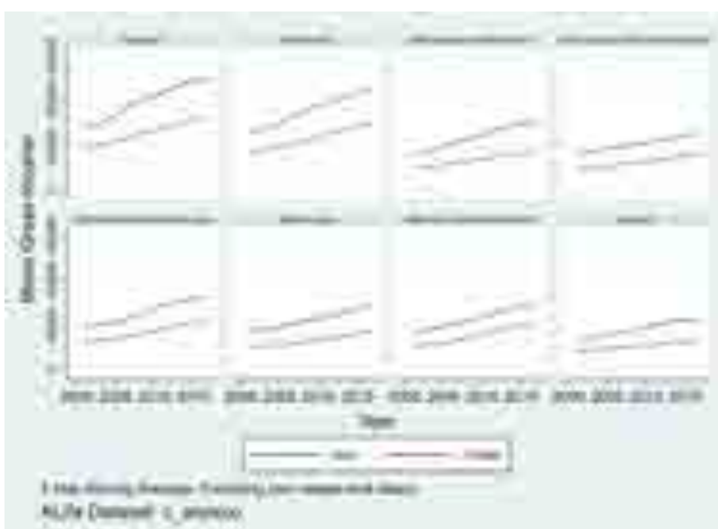
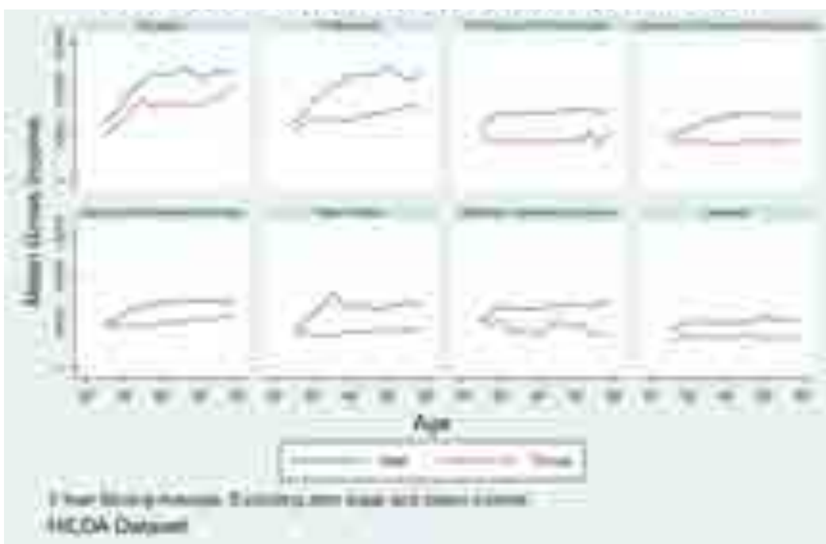
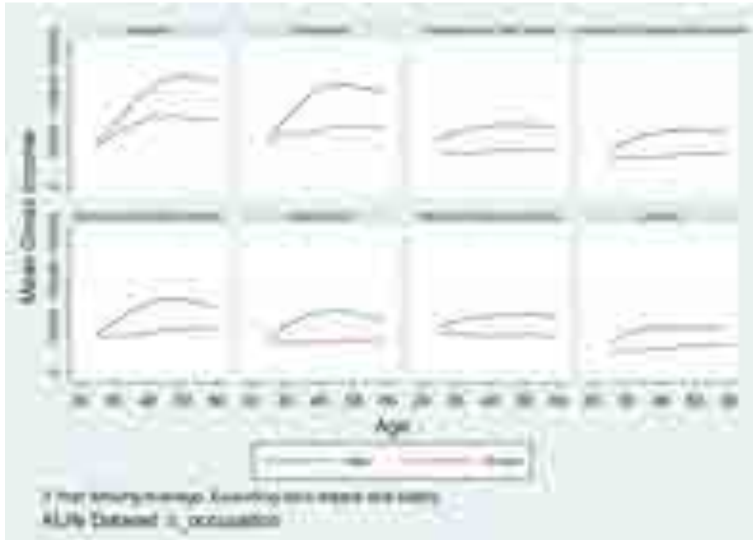


Figure 15 examines gross income across the age distribution split by occupation. In these figures, it is interesting how constant gross income is across age in most professions. This is probably mostly driven by economy-wide increases in wages. Men in managerial and professional roles experience the most dramatic increases in earnings as they age. But again, men consistently earn more than women across all professions and women do not experience as much of an increase in income as they age. This is consistent across the two datasets. It is also consistent with work patterns (full-time/part-time) and differences in labour market experience between men and women across the life cycle.

Figure 15: Mean gross income by occupation type and age





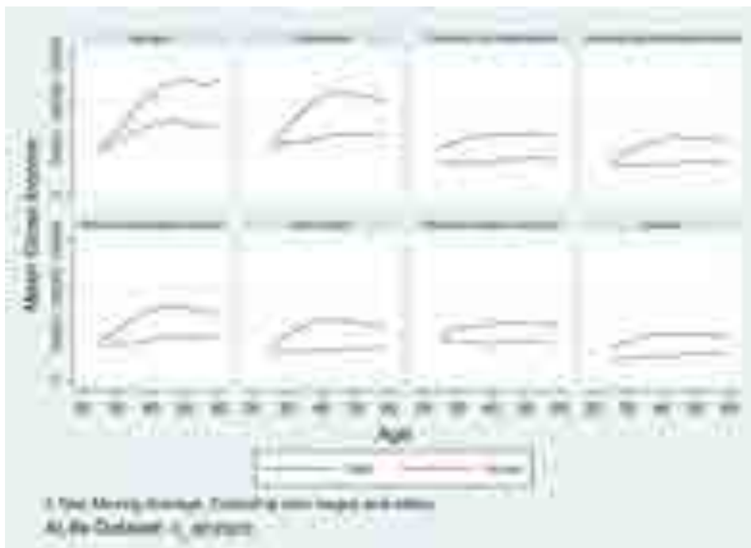


Figure 16 presents a similar picture to Figure 14 in the patterns over time by occupation in the evolution of wages and salary. We again see a parallel, steady increase of wages and salary over time that is constant among men and women. And once again, men on average have higher wages and salary than women. We can also see that wages appear to have grown at a slightly higher rate for men.

Figure 16: Mean wage and salary income by occupation type

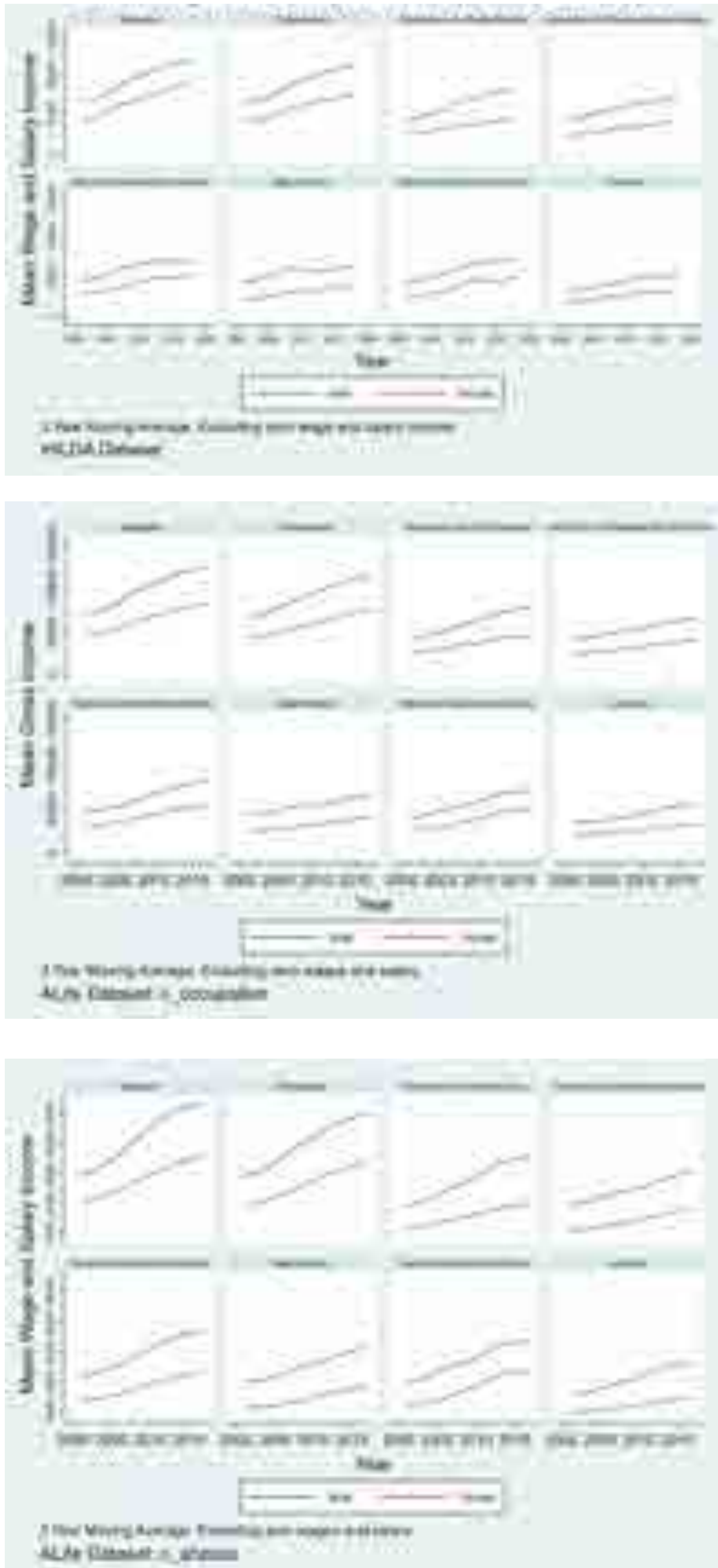
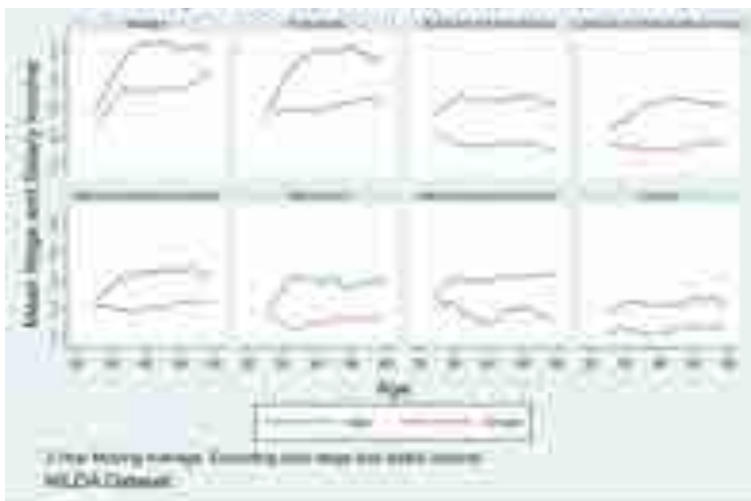
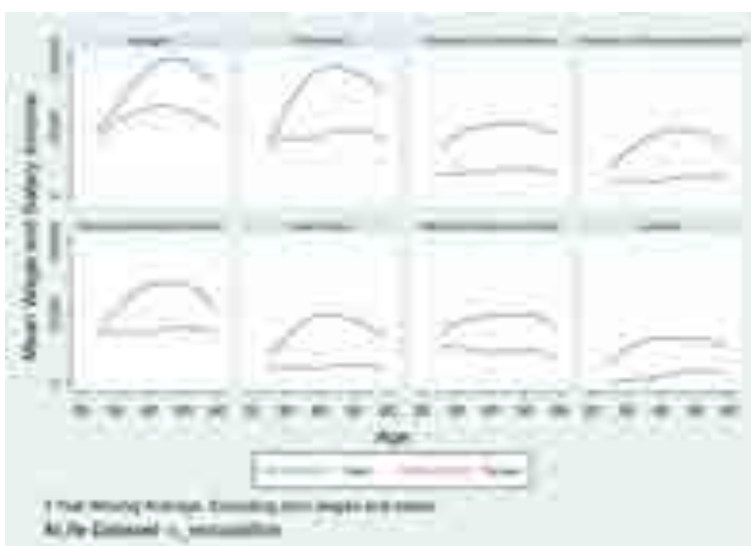
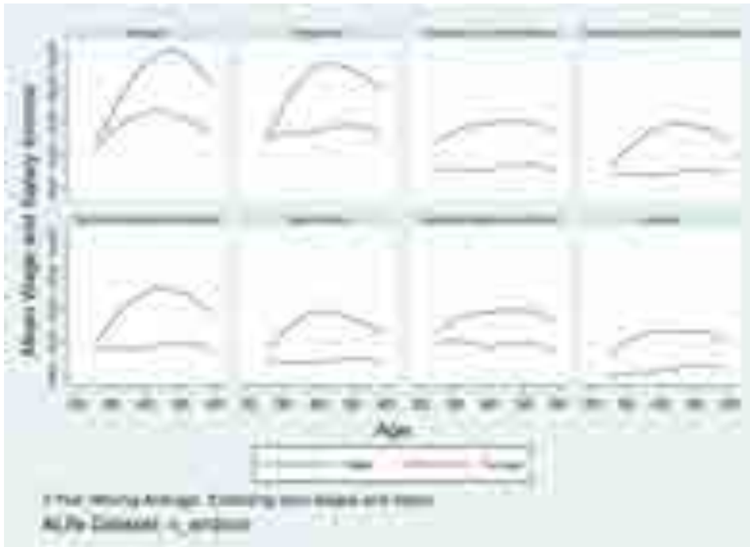


Figure 17 presents the pattern of wage and salary income across different ages split by occupation type. While wages appear to be very flat for women across the age distribution for most occupations, there appears to be a parabolic wage curve for men as they age in the ALife sample. This curve also appears for women in manager roles. This quadratic, inverted-U shaped age-earnings profile is commonly found in labour market studies. In the HILDA dataset, we see a similar trend, but the data exhibits a lot of noise. As with gross income, wage and salary income show the most growth by age for men in manager and professional roles.

Figure 17: Mean wage and salary income by occupation type and age







VI. Summary and discussion

The most striking result from our comparison of occupation in HILDA and the ALife data is the very large difference in reported occupation change in ALife relative to HILDA. This is not surprising, as occupation is not a variable which the ATO requires to assess tax liability. Individual tax-filers have little incentive to update their occupation and tax authorities have little incentive to compel taxpayers to report their current occupation.

The big spike in reported occupation change in ALife during the Global Financial Crisis is difficult to explain. It may be that disruptive job change may induce people to report occupational changes more than in normal times. However, if this were the explanation, we would expect to see this in other data sources such as HILDA. A more likely culprit is some type of administrative change in the ATO systems, as discussed above.

Interestingly, men seem to report more occupation changes in HILDA than women but there is no large difference in the taxation data. Whether women over-report occupational change in the tax data or men under-report in the tax data is hard to determine with our data. It could be that women are more diligent in updating their information with the taxation office.

When looking at the evolution of the occupation distribution over time, we can see important shifts that take place during the twenty years of our data. For men, professions such as managers and professionals have increased at the expense of technicians, trade workers and labourers. For women, we also see increases in the proportion of professionals and managers and decreases in clerical workers.

For both men and women, the top income quintile is dominated by managers and professionals. At the bottom income quintiles, labourers and technicians are the most common occupations amongst men and clerical and community and personal service occupations are most common among women.

Occupational change is lower at older ages and lower at higher incomes. This is consistent with other research and the patterns are very similar for men and women.

There is a lot of variation in occupation change across occupations with some occupations (such as managers) experiencing a lot of change and others (professionals; community and personal service workers) experiencing very little. Women seem to change occupations more often in the typically 'male' occupations and similarly for men who experience more occupation change if they work in typically 'female' occupations. This does not bode well for attempts to reduce occupational segregation.

Finally, we see higher wage growth across age for men in the data than for women, consistent with the gender wage gap and gender patterns in labour force participation in Australia where it is very common for women to work part-time and to take breaks from the labour market. This wage growth with age is particularly prominent for managers.

Understanding the distribution of occupation across different demographic groups can help us understand important issues such as income inequality and the wage gap. For example, people often point to occupational choice as a major component of the wage gap as women tend to choose (or be tracked towards) less lucrative occupations than their male counterparts. While this hypothesis is debated in the literature, our analysis shows a large and continuing difference in the occupational composition of the male and female workforce in Australia.

The Australian Taxation Office is consider releasing a data file which would include household links and which would allow for inter-generational analysis. Other research has found that an individual's choice of occupation and education is highly influenced by their parent's career choice, social network, wealth, and income. Such a dataset would allow for inter-generational analysis of occupational choice that could shed additional light on the occupational distribution in Australia.

Higher occupational mobility has been linked in previous research to higher degrees of growth and productivity, and consequently with higher wage growth. A healthy labour market features a labour force that is continually moving toward more productive and growing industries and occupations. Understanding these trends and being able to track how they respond to policy in different datasets is important.

The findings from this paper will be of interest to the Australian Taxation Office to the degree that the ATO uses this variable in their duties administering Australia's tax system. For researchers, our results will help them to gain a better understanding of the usefulness of the occupation data in ALife. Our analysis suggests that research using ALife which exploits cross-sectional variation in occupation by income, sex or age should, for the most part, produce analysis that is representative of Australia. Analysis that relies on changes of occupation over time, such as fixed effects analysis, will suffer from reduced variation in occupation and may be influenced by biases in which types of occupational changes are captured and which are not. If the unobserved factors which impact on reporting occupational change in ALife are correlated with unobservables which influence outcomes of interest, such analysis may be problematic. While it might be interesting to speculate about the factors which determine under-reporting in ALife relative to HILDA, we are unable to study this question without some way to determine which people who don't report occupational changes actually change occupation. This is fundamentally unobservable in the data.

References

Abhayaratna, T., A. Carter and S. Johnson (2021). 'The ATO Longitudinal Information Files (*ALife*): Individuals – A New Dataset for Public Policy Research'. Tax and Transfer Policy Institute Working paper, forthcoming.

Australian Bureau of Statistics (2006). Australian and New Zealand Standard Classification of Occupations (ANZSCO), First Edition, Australian Bureau of Statistics, Canberra, Cat No. 1220.0

Deutscher, N. (2019). 'Job-to-job transitions and the wages of Australian workers.' Australian Treasury Working Paper 2019-07. Available at: https://treasury.gov.au/sites/default/files/2019-11/p2019-37418-jobswitching-v2.pdf.

Kambourov, G. and I. Manovskii (2009). 'Occupational Mobility and Wage Inequality. The Review of Economic Studies, 76(2): 731-759.

Moscarini, G. and F. Vella (2008). 'Occupational Mobility and the Business Cycle'. National Bureau of Economic Research Working Paper 13819.

Sicherman, N. (1990). 'Education and Occupational Mobility'. Economics of Education Review, 9(2): 163-179.

Sinha, R. (2016). 'Intergenerational Occupational Mobility and Labor Productivity'. The World Bank. Available at SSRN: https://ssrn.com/abstract=2952882

Summerfield, M., S. Bright, M. Hahn, N. La, N. Macalalad, N. Watson, R. Wilkins and M. Wooden (2019). 'HILDA User Manual—Release 18'. Melbourne Institute: Applied Economic and Social Research, University of Melbourne.

Watson, N., and Summerfield, M. (2009), 'Assessing the Quality of the Occupation and Industry Coding in the HILDA Survey', HILDA Project Discussion Paper Series No. 3/09, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N. and M. Wooden (2010). 'The HILDA survey: progress and future developments'. Australian Economic Review, 43: 326–336.

Appendix

Results including observations with zero salary and wages

In this appendix we present versions of the main results in the paper which include those observations where people report zero wages and salary. In the ALife data, this was only around one per cent of the data. In HILDA, this was closer to 10 per cent. As a result, the results for ALife are almost entirely unchanged. For the HILDA data, some of the numbers change but none of the main conclusions that we draw from the examination of the two datasets are affected. For Figures A1 – A4 and A9 – A10 we show the comparison between occupation in HILDA and the "c_occupation" variable in ALife. As the results for ALife change so little, we do not show the "c_anzsco" variable.

Table A1: Percent Occupation Change in 2017
(including zero wages and salary—compare to Table 5)

Variable	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Total Change	19.0%	5.4%	7.1%
Female	17.3%	5.6%	7.1%
Male	20.5%	5.1%	7.1%
Sample size	7,231	559,546	757,331

HILDA wave 17 (2017); ALife2017 (2016-2107 financial year)

Table A2: Percent Occupation Change across all years
(including zero wages and salary—compare to Table 6)

Variable	HILDA	ALIFE c_anzsco	ALIFE c_occupation
Total Change	21.8%	7.4%	9.7%
Female	19.7%	7.5%	9.5%
Male	23.6%	7.3%	9.9%
Sample size	104,258	7,423,559	10,080,683

HILDA wave 17 (2017); ALife2017 (2016-2107 financial year)

Table A3: Demographic Variables in 2017
(including zero wages and salary—compare to Table 9)

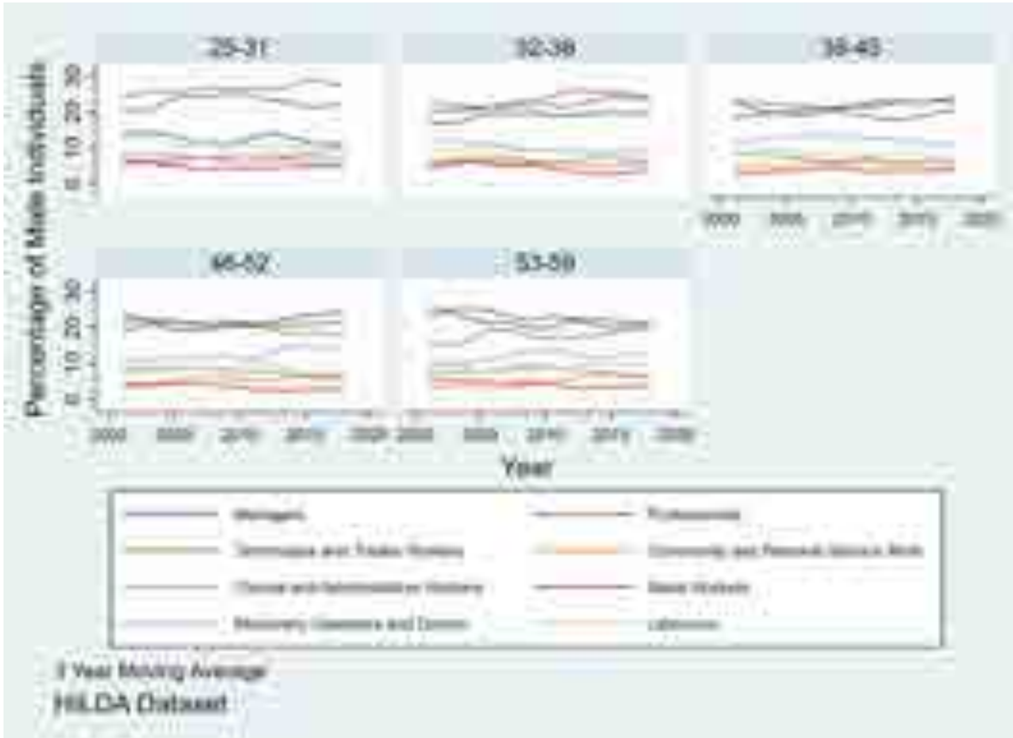
Variable	HILDA	ALIFE
Percent Male	52.9%	51.1%
Percent Female	47.1%	48.9%
Mean Age	41.2	40.6
Percent Married	58.0%	59.8%
Percent Partnered	73.6%	n/a
Mean # of Dependents	0.74	0.92
Mean Gross Income	$82,112	$74,087
Mean Wage and Salary Income	$69,282	$66,108
Sample size	8,063	

HILDA wave 17 (2017); ALife2017 (2016-2107 financial year)
HILDA data weighted using cross-sectional, responding person population weights

Table A4: Geographic Variables in 2017
(including zero wages and salary—compare to Table 11)

Geographic Area	HILDA	ALIFE
NSW	31.8%	31.9%
VIC	27.2%	25.6%
QLD	19.3%	19.6%
SA	6.6%	6.6%
WA	10.1%	10.9%
TAS	2.1%	2.0%
NT	1.0%	1.1%
ACT	1.8%	2.0%
Other	0.0%	0.0%
Major City	71.1%	75.6%
Inner Regional	20.0%	15.8%
Outer Regional	7.8%	7.3%
Remote Area	1.1%	1.0%
Very Remote Area	0.1%	0.3%
Urban Area	84.6%	89.0%
Not Urban	15.3%	11.0%
Sample sizes	8,063	842,457

Figure A1: Distribution of occupation by age group: males (compare to Figure 2)



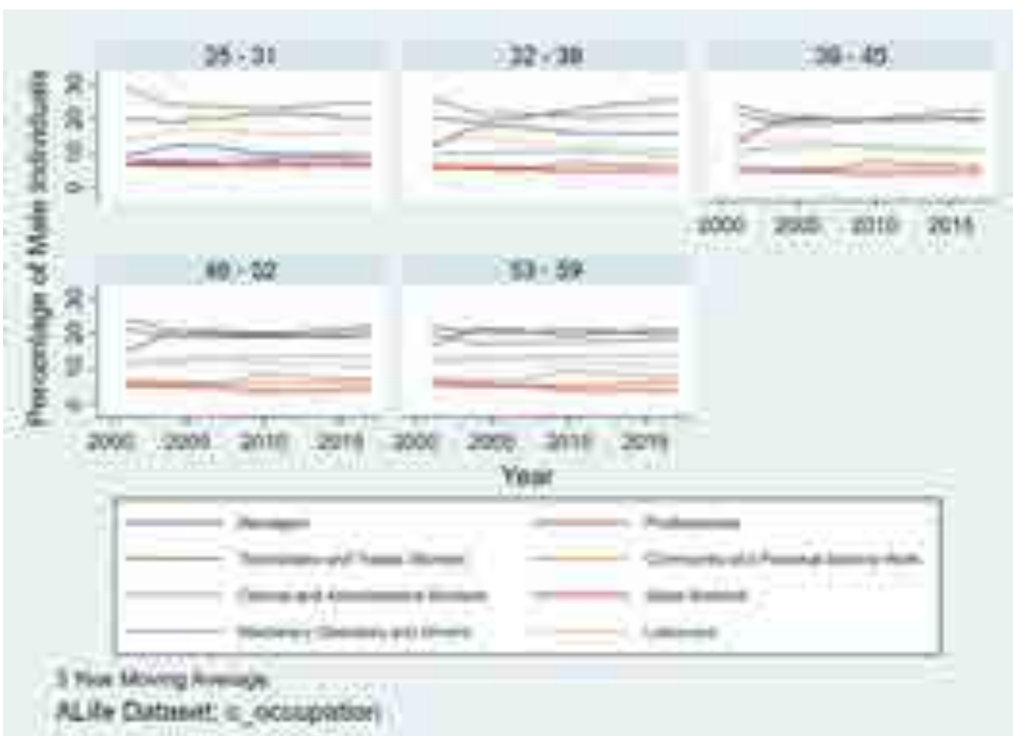
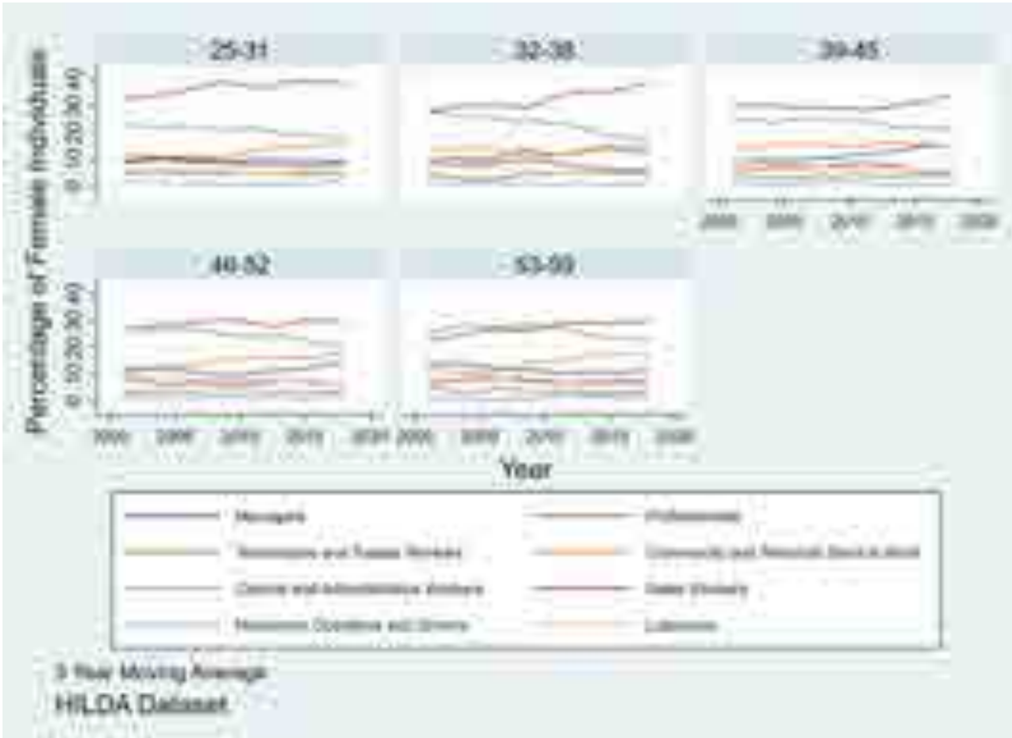


Figure A2: Distribution of occupation by age group: females (compare to Figure 3)



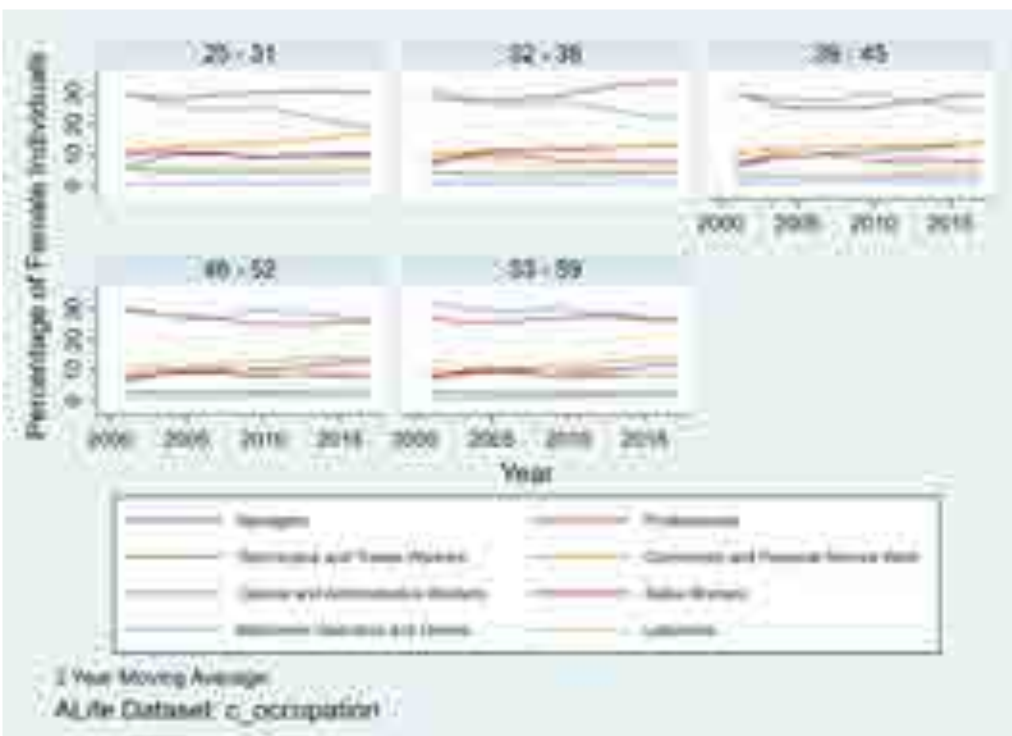
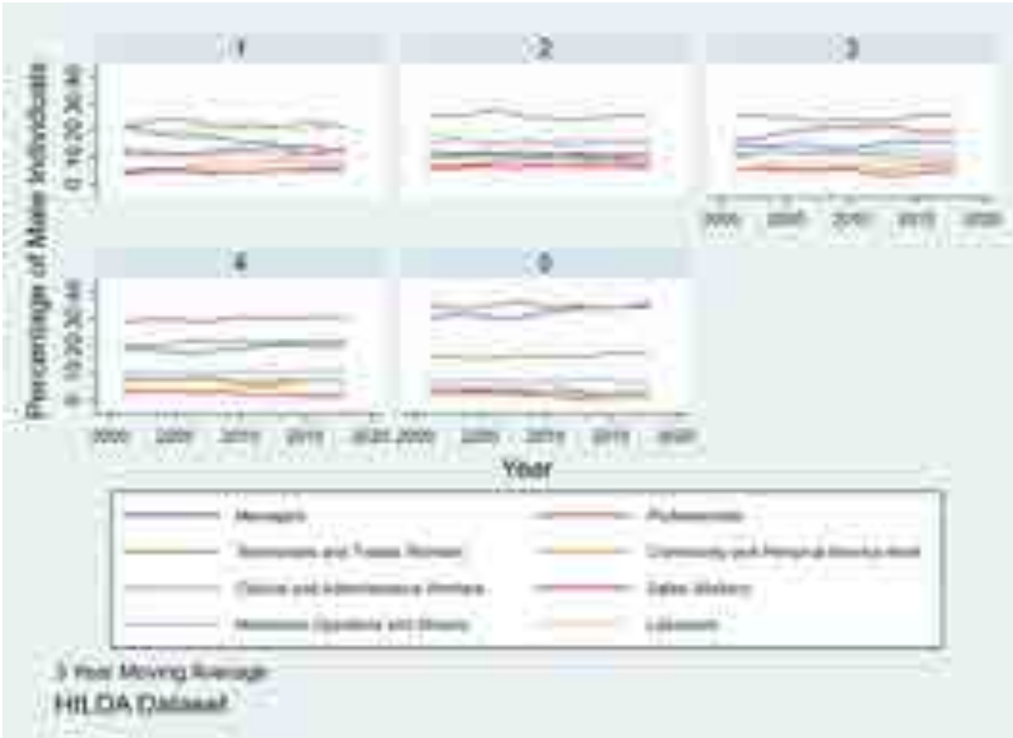


Figure A3: Distribution of Occupation by Gross Income Quintile: males (compare to Figure 4)



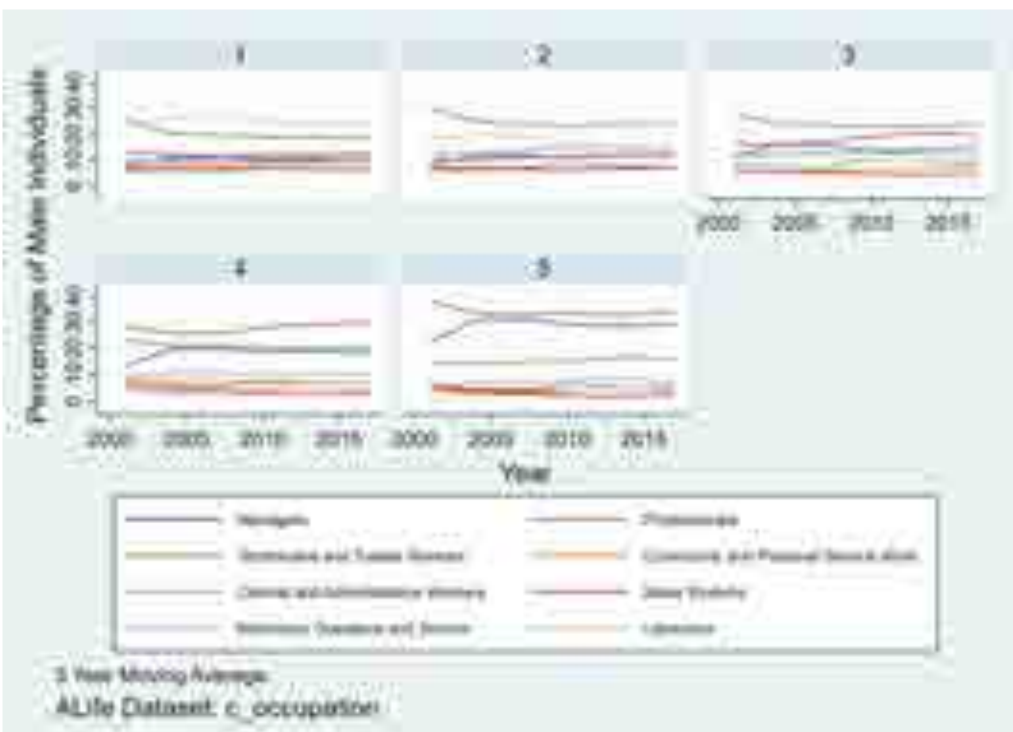
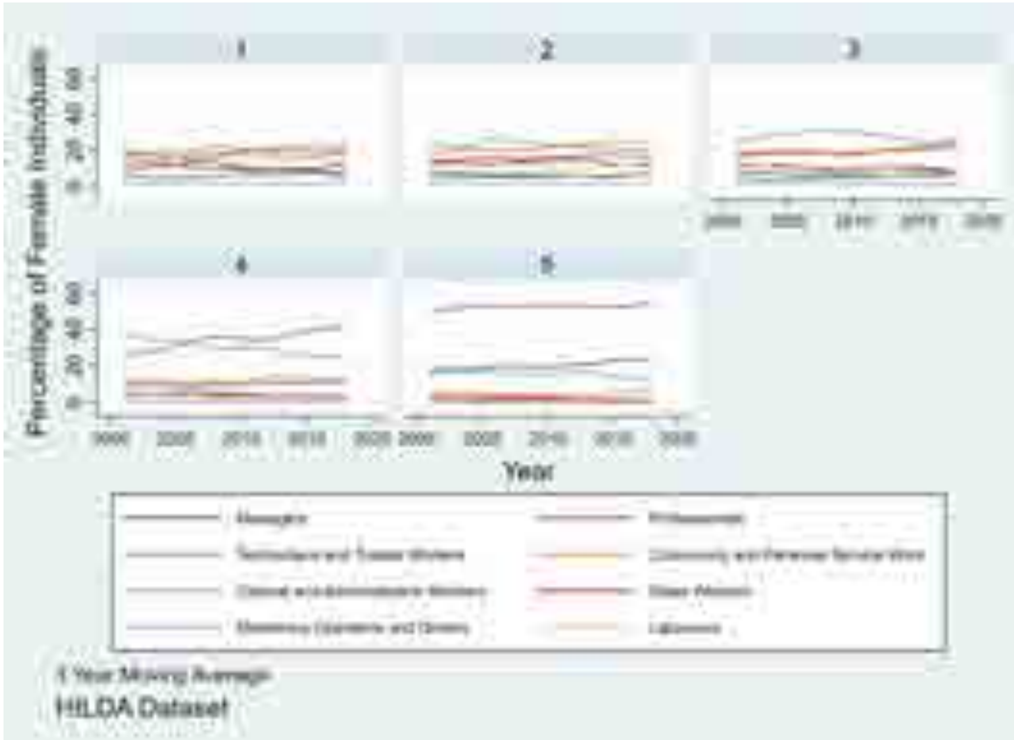


Figure A4: Distribution of Occupation by Gross Income Quintile: Females (compare to Figure 5)



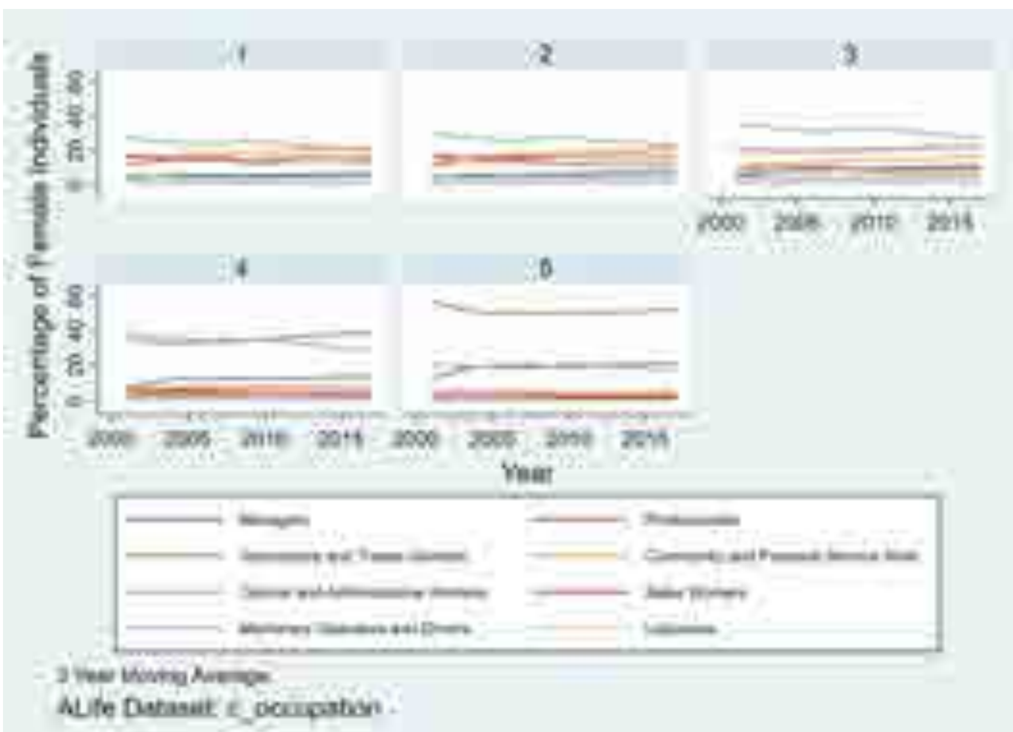
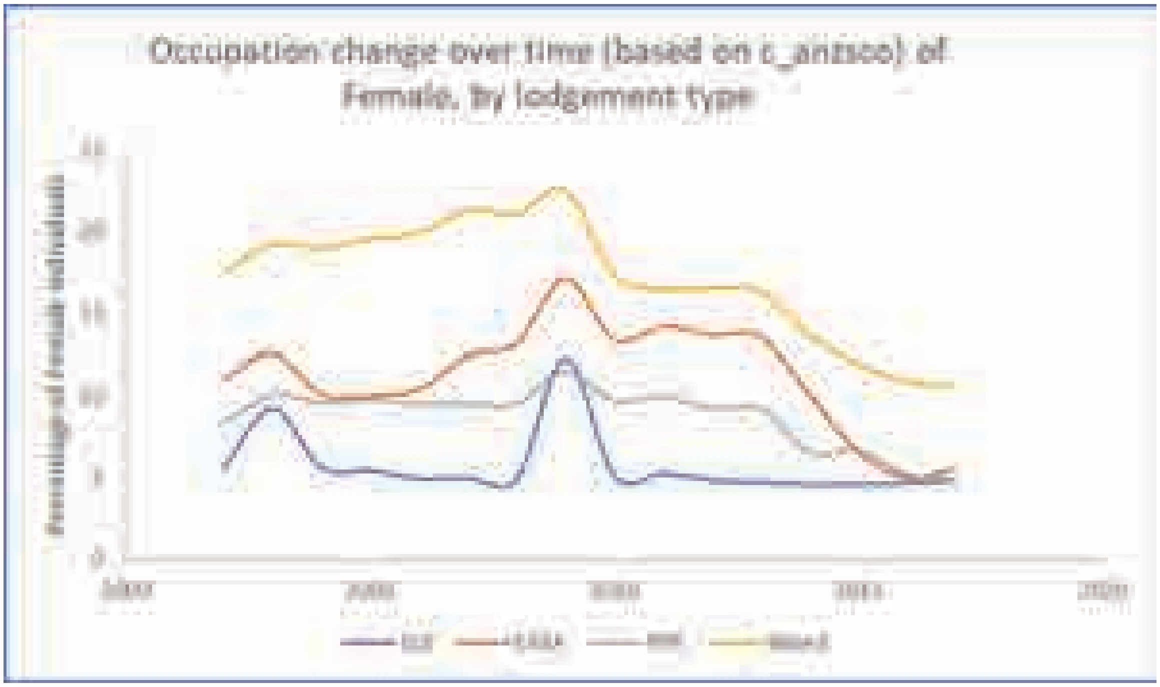
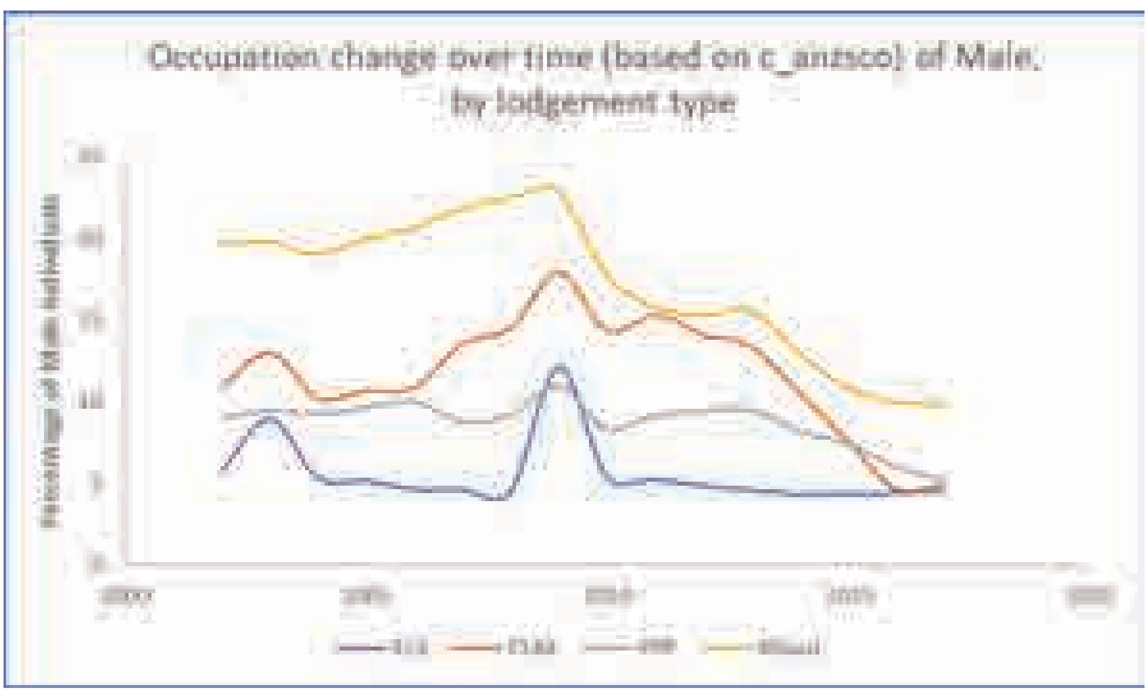
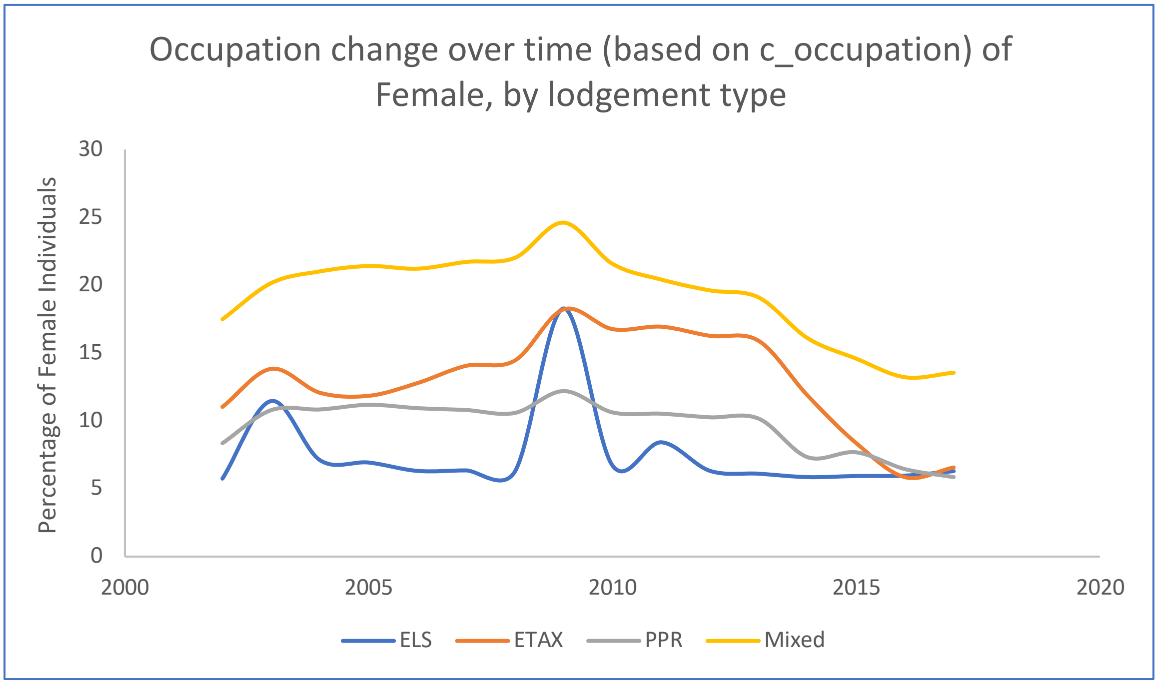


Figure A5 – A8: Occupation Change Over Time, Split by Lodgement Type and Sex

(compare to Figure 8)






Occupation change over time (based on c_occupation) of Female, by lodgement type
Percentage of Female Individuals
30
25
20
15
10
5
0
2000
2005
2010
2015
2020
ELS
ETAX
PPR
Mixed

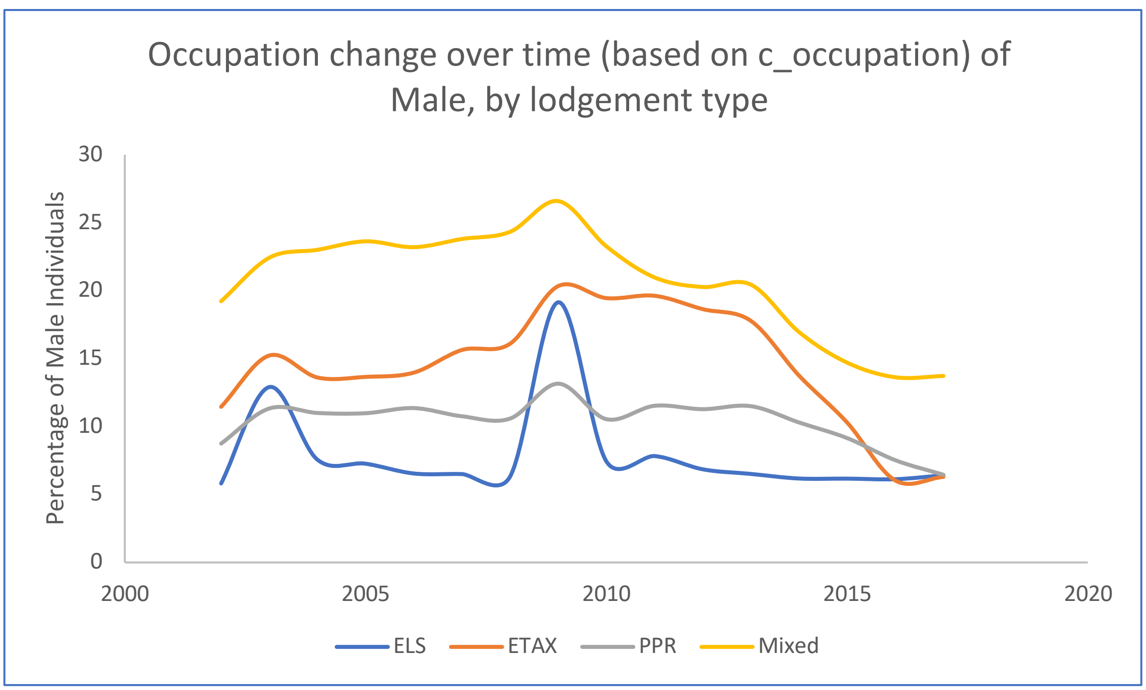

Occupation change over time (based on c_occupation) of Male, by lodgement type
Percentage of Male Individuals
30
25
20
15
10
5
0
2000
2005
2010
2015
2020
ELS
ETAX
PPR
Mixed

Figure A9: Occupation Change by Age Group (compare to Figure 10)

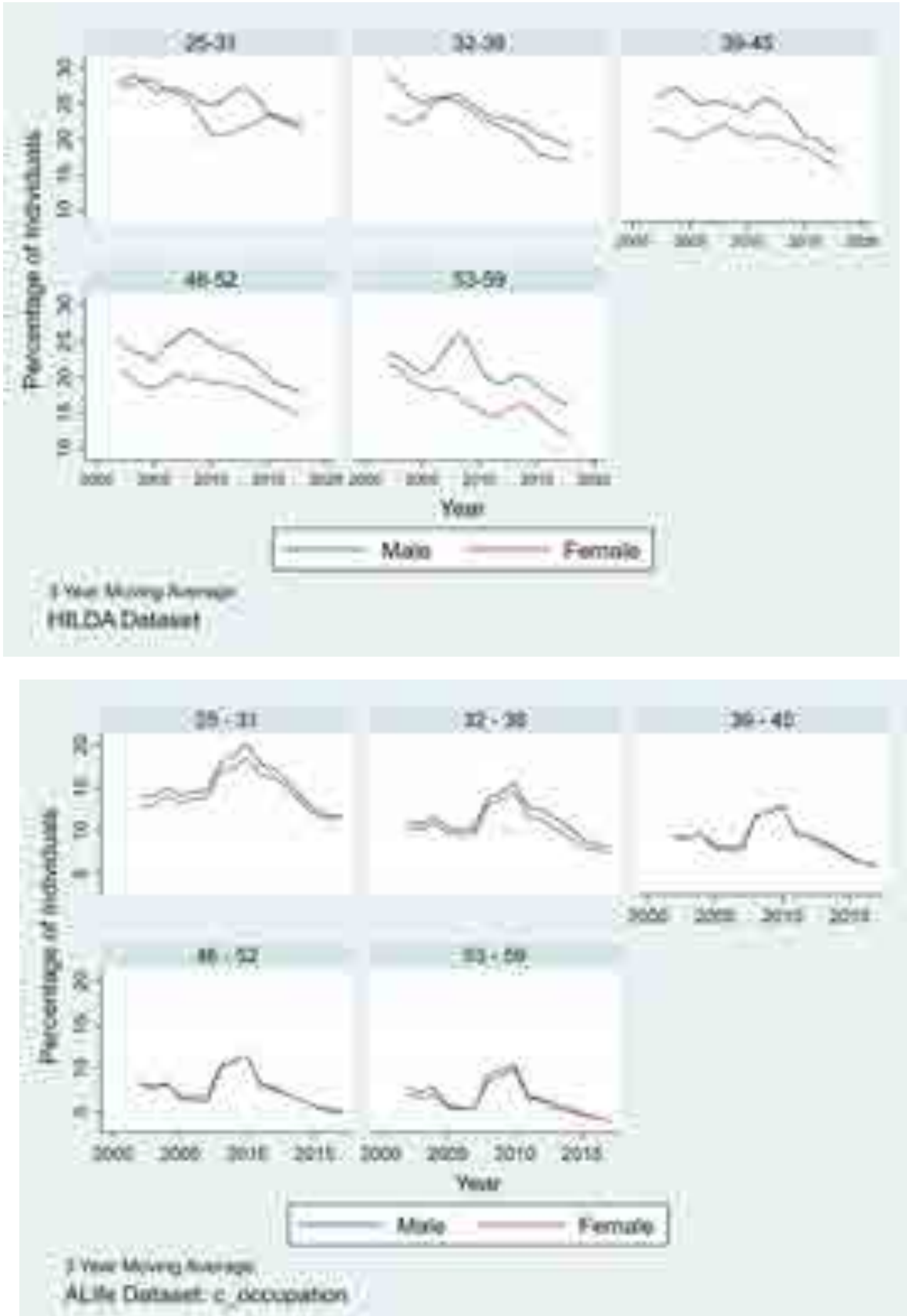


Figure A10: Occupation change by gross income quintile (compare to Figure 11)