**Economics Working Paper Series**

*Faculty of Business, Economics and Law, AUT*

# Human Capital Formation and Changes in Low Pay Persistence

Kabir Dasgupta and Alexander Plum

2020/15

# Human capital formation and changes in low pay persistence

Kabir Dasgupta, Alexander Plum*

**Abstract**

This study aims at understanding how persistence in low pay changes over time. In particular, we extend the existing literature on human capital formation by documenting heterogeneity in low pay persistence by age and human capital level. We utilise population-wide tax records to track monthly labour market trajectories of workers who are observed in low paid employment during the initial period of analysis. Performing age- and qualification-specific regressions, our empirical findings indicate that low pay persistence reduces with time. However, the magnitude is highly heterogeneous across the workforce. For a qualified worker in their early 20s, the risk of staying on low-pay declines by, on average, 5 to 10% points after one year—while for a worker in their 50s, independent of their qualification level, persistence remains almost unchanged. We find a strong association between decline in low-pay persistence and the firm's average wage level.

**JEL Classification**: Low pay, human capital formation, state dependence, random-effects probit , initial condition , unobserved heterogeneity
**Keywords**: J63, J61, J24

*Kabir Dasgupta, Auckland University of Technology (AUT), Email: kabir.dasgupta@aut.ac.nz; Alexander Plum, Auckland University of Technology (AUT), Email: alexander.plum@aut.ac.nz. All errors are our own.

# 1 Introduction

The life-cycle wage growth pattern has been widely studied in economic literature. Its well-documented hump-shape suggests that an individual's wage growth rate declines over their life-cycle (e.g., Low et al., 2010; Lagakos et al., 2018). Moreover, empirical evidence supports a positive correlation between an individual's wage growth and their level of human capital. These findings have important implications for the identification of low-pay employment across a workforce, both at a particular time point and with respect to changes in low-pay risk over time. First, the share of low-pay employment is skewed towards younger workers.[1] Second, the qualification composition of those on low pay can be age-dependent with a larger fraction of qualified workers belonging to younger age groups. Third, the risk of staying on low pay over time declines more strongly for younger workers because they demonstrate rising productivity levels (Farber and Gibbons, 1996) and improve employer matches (Topel and Ward, 1992; Abowd et al., 1999) which is likely to accelerate wage growth rates. Our study explores empirical evidence pertaining to the last implication in particular.

Stewart and Swaffield (1999) noted that 'low pay has become an increasingly important policy issue' and that the 'extent of the persistence in low pay have important policy implications' [p. 23]. From a policy-maker's perspective, understanding the transience of low-paid employment and identifying effective mechanisms for entry into higher-paid employment is crucial for efficiency in the labour

[1]For example, Metcalf (1999, p. F49) noted: the "incidence of low pay is far higher among 18 to 20 year olds (...) than those aged 21+".

market and improving social welfare. For example, if low-pay persistence is exceeded by the probability of entering higher-pay, then ‘any job is better than none’ and barriers that hinder job creation (e.g., a high level of minimum wage) should be lifted. Conversely, if “workers became trapped in low-paid jobs the implications would be potentially more serious” (Sloane and Theodossiou, 1996, p. 657) and labour market policies like programs for skills development might be crucial to improve workers’ potential to transit into higher pay jobs.

The number of research paper estimating labour market dynamics of low-paid workers has increased over time (e.g., Plum, 2019; Cai et al., 2018; Fok et al., 2015; Mosthaf, 2014; Clark and Kanellopoulos, 2013; Stewart, 2007; Stewart and Swaffield, 1999). The common approach to quantify the intertemporal relationship of low-pay employment is to employ ‘dynamic/transition models that include both a lagged response and a random intercept’ (Skrondal and Rabe-Hesketh, 2014, p. 211). This type of model has found widespread attention in the econometric literature to identify state dependence, for example, with respect to unemployment (Arulampalam et al., 2000), poverty (Ayllón, 2015; Biewen, 2009), benefit recipient (Bhuller et al., 2017; Wunder and Riphahn, 2014), health (Haan and Myck, 2009), sovereign credit ratings (Dimitrakopoulos and Kolossiatis, 2016), and financial asset holdings (Alessie et al., 2004). The underlying thought is that the lagged dependent variable has a *genuine* impact on the outcome variable (Heckman, 1981a). Although several studies acknowledge that state dependence in low pay may be heterogeneous across the workforce (for instance, Plum, 2019; Fok et al., 2015), to the best of our knowledge, no study so far has accounted for changes in low-pay persistence over time caused by differences in wage growth. To identify time trends in low pay persistence, we use New Zealand's Integrated Data Infrastructure (IDI) to track a sample of workers who were in low-pay employment during the initial period of our analysis (January to March 2013) and are continuously employed over a time window of three years until March 2016. Focusing on month-to-month transitions into higher-paid (or low-paid) employment, we determine how state dependence in low-pay employment evolves over time. Keeping in mind that most young workers face higher wage growth rates, we expect that the chances of staying on low pay evolve heterogeneously over time, depending on age and human capital level. We differentiate between three qualification categories (no qualification, low qualification and high qualification) and perform separate regressions for each age (in years) cohort ranging from 20 through 60. We account for time-trends in low-pay persistence, which contrasts to existing studies that have assumed that the probability of staying on low pay is constant over time. The results indicate that low-pay persistence drops most markedly for young workers aged 20-25 and who have some level of education (indicated by either low or high qualification). For example, the probability of staying on low pay after one year drops by about 9% points (or 5-6% points) for highly (or somewhat) qualified workers aged between 20 and 25. However, for their fellow age cohort members without any qualifications, the estimated decline hovers around 1-2% points. On the other end of the age spectrum (50+), we find that low-pay persistence hardly changes with time and there are almost no differences between the three qualification levels.

Low et al. (2010) explain that much of wage growth over the life cycle "is due to search leading to improved matches" [p. 1453]. This is backed by our data, showing that within the three-year-period, young low-paid workers move into better-paying firms. Moreover, when looking at their earnings level five years later, we find that of those workers in their early 20s, only 30 per cent are still on low pay—while the respective share is two times higher for workers in their 50s. However, the differences between the qualification levels are stark. In particular, workers with some level of educational qualification are much more likely to transit into higher paid jobs compared to when holding no qualification.

Our analysis reveals that modelling low-pay persistence must be undertaken with caution, because changes that occur over time need to be taken into account and such changes are heterogeneous across age and qualification levels. While being in low-paid employment appears to be a temporary labour market status for a sizeable share of younger workers who are at the starting point of their professional career, we do observe substantial low pay persistence for individuals without any qualification as well as for older workers. As such our empirical analysis questions the efficacy of public policies that are often implemented in a 'one-size-fits-all' format or more precisely without considering significant demographic heterogeneity in the population.

The remainder of our paper is structured as follows: Section 2 discusses the literature on human capital formation and low pay, Section 3 describes the data used and its descriptive statistics, Section 4 outlines the empirical identification strategies, Section 5 presents the results, and Section 6 concludes.

# 2 Literature Review

## 2.1 Human capital formation and wage growth

It has been well-documented in the empirical literature that, over the life-cycle, wages follow a concave pattern when plotted against age. This pattern can be explained by firm-specific wage growth that occurs during periods of on-the-job training (Brown, 1989) and improved employer-employee matches (Topel and Ward, 1992; Abowd et al., 1999). The magnitude by which wages grow also positively depends on the individuals' qualifications (Low et al., 2010; Lagakos et al., 2018). This has three important implications when determining persistence in the context of low pay literature:

1. The likelihood of being in a low-paid job is higher for younger workers,

2. Qualification level among low-paid workers is heterogeneous across age cohorts, with a larger proportion of highly qualified low-paid workers belonging to younger age groups,

3. Younger low-paid workers are more likely to exit the low-pay sector, although this likelihood varies by qualification levels.

Another factor that helps in understanding the differences in earning transitions between cohorts is described by the theory of labour market signalling. McCormick (1990) argues that skilled workers prefer utilising on-the-job-search to move between employment. This is because accepting interim low-skilled jobs might send out a negative signal about the true productivity of a worker, which is mostly unknown to the employer. This may eventually lead to positive assortative sorting, where productive workers end up in better-paying firms (supportive empirical evidence has been found by, e.g., Mendes et al., 2010; Abowd et al., 1999). However, for a young worker, such signaling ability might be obscured by their lack of labour market experience. Additionally, since finding an appropriate employee-employer match might be a time-consuming process, signaling can play a relatively more prominent role for more experienced workers' labour market endeavours.

## 2.2 Low-pay dynamics

Our study investigates how the chances of exiting low-pay change with time. The effect of low-pay employment on an individual's future labour market status has gained substantial attention in the past. The majority of the studies compare the labour market dynamics of the low-paid with the unemployed to detect 'no-pay–low-pay' patterns, or a springboard effect (e.g., Cai et al., 2018; Stewart, 2007; Stewart and Swaffield, 1999). In recent years, the empirical literature has adopted a standard approach in estimating state dependence in low pay. Represented by a first-order Markov chain, the basic concept is to include (in most studies by one-period) lagged labour market position on the right-hand side of the equation to estimate persistence in a labour market status. As individuals are likely to differ in their unobservable characteristics (Heckman, 1981a), not controlling for individual-specific time-invariant effects may lead to an overestimation of state dependence (Stewart, 2007).

When looking at the prospects of moving from low pay into higher pay, a certain degree of permeability is found. For example, Sloane and Theodossiou (1996, p. 665) found that "only 44.4% of the low-paid in 1991 remained in this category 2 years later", which leads to the conclusion that "low pay is a temporary phenomenon". These findings have been backed by numerous studies, including Cai et al. (2018); Mosthaf (2014); Clark and Kanellopoulos (2013) who find that the probability of staying on low-pay is exceeded by the chances of entering higher pay.

To our knowledge, not much attention has been paid to a detailed understanding of heterogeneity in labour market dynamics (e.g. low-pay persistence) across age groups. At best, a few studies have tried to account for age-specific differences by controlling for age indicators in their empirical models. For example, in his study on UK's low-wage dynamics, Stewart (2007) uses the British Household Panel Survey (BHPS) and accounts for the entire working age by including men and women of ages between 18 and the respective state retirement threshold (which is 65 for men and 60 for women). This approach was later adopted by Cai et al. (2018). Similarly, using the Survey on Households Income and Wealth from Italy, Cappellari (2007) analyses a sample that incorporates male workers aged 18-60 and female workers aged 18-55. Furthermore, using the first 11 survey waves of the Household, Income and Labour Dynamics in Australia (HILDA), Fok et al. (2015, p. 877) restrict their sample to individuals aged 21–54 years "because of the potential complications arising for persons transitioning from study to employment. Likewise, we also omit persons aged 55 years and older due to the transition between employment and retirement." Also using HILDA data, Buddelmeyer et al. (2010) further drop individuals below 21 from the population of interest. A higher minimum age bar was set by Mosthaf (2014), who only allows for individuals older than 29 to be part of the sample. Evidently, the literature has not found any consensus on the particular age window to focus on while evaluating labour market dynamics. This is likely due to spatial differences in a state's definition of working age population.

Additionally, as already highlighted, qualification of the low-paid worker might be heterogeneously distributed across age cohorts. The literature on low pay has identified qualification as a vital indicator of human capital that could be utilized to understand dynamics into and out of low pay. For example, Cappellari (2007) finds that a higher level of human capital reduces the risk of entering low pay but elevates the chances of exiting the low pay sector only marginally. Also Mosthaf (2014) find that higher level of qualification reduces the risk of entering low pay from higher pay and improves the chances of transitioning from low pay into higher pay, although most differences are not statistically significant different from each other. Indications on the effect of human capital provided by Plum (2019) shows that a low-paid worker in a higher-skilled occupation has a significantly higher chance of entering higher-paid employment compared to low-paid worker in a low-skilled occupation. While these studies have focused on the interrelation of qualification and earnings prospects of low-paid workers, the vast majority of studies has included qualification merely as a covariate.

A study worth mentioning in this context is Fok et al. (2015), which explicitly allows for heterogeneity in the effect of the lagged labour market position. For this reason, the authors "add interactions between variables for various demographic characteristics and the lagged labour force status variables" [p. 886], including age and educational attainment. The authors find that the state dependence in low pay is larger for those in older age groups. For example, "for men, compared with 21–29 year-old's, low-paid employment increases the probability of remaining low paid in the next year by 4 percentage points for 30–39 year old's (11.51-7.51) and by 4.9 percentage points for 40–54 year old's (12.36-7.51)" (Fok et al., 2015, p. 890). Differences are also found with respect to educational attainment, and the findings indicate that low-pay persistence is higher for people with higher educational attainment. To sum up the current literature on low pay, although recent studies have acknowledged the heterogeneous aspects of exiting low pay across age and human capital, the empirical specifications have assumed that persistence in low pay is constant over time and as such, have not accounted for time trends. However, Cai (2019) replicates the study of Fok et al. (2015) and finds that when accounting for correlation in unobserved heterogeneity, "there is no evidence on heterogeneity in the low-pay no-pay cycle across the demographic subgroups" (p. 1493).

# 3 Data and Descriptive Statistics

For our empirical investigation we use administrative data from Statistics NZ's Integrated Data Infrastructure (IDI). The IDI contains population-wide longitudinal microdata about individuals, households, and business enterprises. These data are sourced from various government and non-government agencies, as well as Statistics NZ surveys. The data are confidentialised by means of assigning a unique identifier to each individual.[2]

To derive our population spine, we start with the 2013 Census which was conducted in March of the same year. The Census holds a range of information on individual and household characteristics, including age, qualification, ethnicity, location, gender, household size, etc. First, we restrict our sample to men who were aged between 20 and 60 (including) in March 2013. Moreover, we trim the sample to those individuals with a full set of information on their characteristics that we use as controls in our regression model. Next, we link these individuals to their tax records from Inland Revenue (IR). IR records seven different income sources (wages and salaries, benefits, paid-parental leave, withholding payment, compensation claims, NZ superannuation, and student allowances) on the monthly level and we match income information from wages and salaries for the period January 2012 to March 2016.[3]

[2]For further details please visit `https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/` and see the Disclaimer in the Appendix.

[3]An employer-identifier enables us to determine the month-specific number of employees of an employer. To exclude self-employed, we drop individuals who received wages and salaries for a minimum of one month from an employer with only one employee on his payroll.

Our focus is on understanding how independent of composition effects of the population the individuals' wage position changes, either caused by productivity increase or by an improved employer-employee match. However, one potential pathway to move from low pay into higher pay is that the income distribution within the population changes as the workforce grows or shrinks. For example, an influx of seasonal worker would increase the number of worker with a small wage. To avoid these spillover effects, we construct a balanced sample by trimming the sample to men who received income from wages and salaries in each month of the period January 2013 to March 2016 (the employment characteristics for 2012 are used as controls in the regression). This leaves us with a balanced sample of 601 686 men (see Table 1). It is worth noting that restricting the sample to continuously employed worker causes a selective sample: the unemployment risk is not equally distributed across age and is elevated for especially younger worker. Accounting for unemployment is useful to uncover whether low pay can act for the unemployed as a 'stepping stone' towards higher pay or whether is makes more sense to wait for a higher-paid employment. However, skills may deteriorate during non-employment spells, which will have an effect on low-pay persistence. To avoid this additional layer of complexity, we chose to restrict the sample to continuously employed.

To approximate the human capital level, we use the information provided in the 2013 Census on the highest qualification an individual achieved.[4] This in-

[4]The relevance of the qualification level is not age-independent. Due to the lack of prolonged labour market experience, the qualification level might be of higher relevance for young workers and for older workers, the occupation might play a more important role. To have a consistent

**Table 1:** Population spine

| | Census 2013 | On low-pay | |
|---|---|---|---|
| | | March 2013 | January to March 2013 |
| $N$ | 601 686 | 120 342 | 26 487 |

formation is structured according to the New Zealand Qualification Framework (NZQF), which is categorized in the Census in the range 0 (no qualification) to 10 (doctoral degrees). For our study, we form the following three qualification groups: no qualification, Level 1 to 4, and Level 5-6 and higher. To put the qualification levels into better perspective, we follow the European Commission (2017) and show their position in terms of the European Qualifications Framework (EQF). NZQF levels 1 to 4 range from lower-secondary education to upper secondary general school-leaving certificates and refer to the EQF levels 2-4. NZQF levels 5-6 are higher professional qualifications and refer to EQF level 5. We do not decompose the higher qualification group further to avoid sample size issues resulting from a small number of low-paid workers. Figure B.1 shows the monthly pattern of income from wages and salaries across age for different qualification groups for March 2013 (see also Table 1). We can clearly see the hump-shaped distribution, though the wage progression below age 40 is especially strong for the workers with higher levels of qualification. Conversely, though some income gain can be detected for workers without any qualifications, their income progression flattens after age 30.

category across all age groups, we chose the highest qualification level as a proxy for human capital.

In the next step, we determine our low-pay threshold. There exist different approaches (a discussion can be found in Fok et al. (2015)) and we follow Cappellari (2007) who looks at the relative position within the earnings distribution. This means that there is a fixed share of low-paid workers in the population and temporal changes (such as macroeconomic influences that have an impact on the earnings but not on the individual's position in the distribution) will not influence our estimation. As cut-off point for low-pay, we define anyone with an aggregated[5] monthly gross earnings that belong to the two lowest deciles as earning low pay and those with earnings above that threshold are defined as higher-paid. To put the threshold into perspective, note that the minimum wage in March 2013 stood at NZ$13.50 for an adult, which results in a monthly wage (taking 20 working days) when working 40h/week of NZ$2160. The respective low-pay threshold is NZ$2936. Furthermore, our findings are not affected by other commonly used low-pay thresholds. This results in a sample of 120 342 individuals (second column of Table 1), and when looking at the low-pay distribution across age, approximately half are below the age of 30. However, for individuals who are observed to be on low pay in March 2013 (the Census month) there might be differences in experiences of being in low-paid employment. For instance, while an individual may be low-paid only the month of March, it might be the case that another worker in our sample has experienced low-pay since well before March. To ensure that the group of low-pay workers have a certain attachment to the low-pay sector, we trim the sample further and include only to those individuals who were

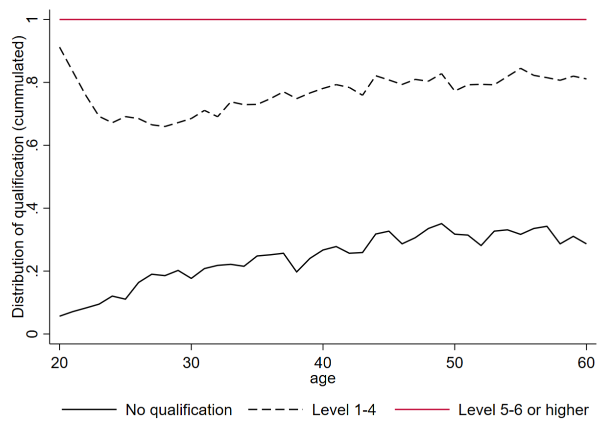
[5] Some individuals either hold multiple jobs per month or transit between two jobs in a month.

on low pay the entirety of January to March 2013. This leaves us with a sample of 26 487 individuals (third column of Table 1). Where similarly to before, every second low-pay worker aged 30 or younger. Also Figure B.2 indicates that the low-pay distribution is strongly skewed towards younger workers.

Note that we cannot observe the numbers of hours an individual is working. Thus, changes in earnings might be caused by a higher wage rate or by extending hours worked, e.g., moving from part time to full time (this aspect is more relevant for women and for this reason we have restricted our analysis to men). This might be more relevant for younger worker as they might work besides studying. In one robustness estimation, we restrict our population spine to worker who state in the 2013 Census to be full-time employed. In a second robustness estimation, we link the individuals with the 2018 Census and restrict the sample to workers wit unchanged qualification level. In a third robustness estimation, we shorten the covered period to March 2015. In all three cases, the findings are hardly affected. A second limitation is that restricting the sample to continuously employed worker causes a selective sample: the unemployment risk is not equally distributed across age and is elevated for especially younger worker. Accounting for unemployment is useful to uncover whether low pay can act for the unemployed as a ‘stepping stone’ towards higher pay. However, skills may deteriorate during non-employment spells, which will have an effect on low-pay persistence. To avoid this additional layer of complexity, we chose to restrict the sample to continuously employed.

To provide further insights into the characteristics of low-paid workers, Fig-

**Figure 1:** Qualification distribution

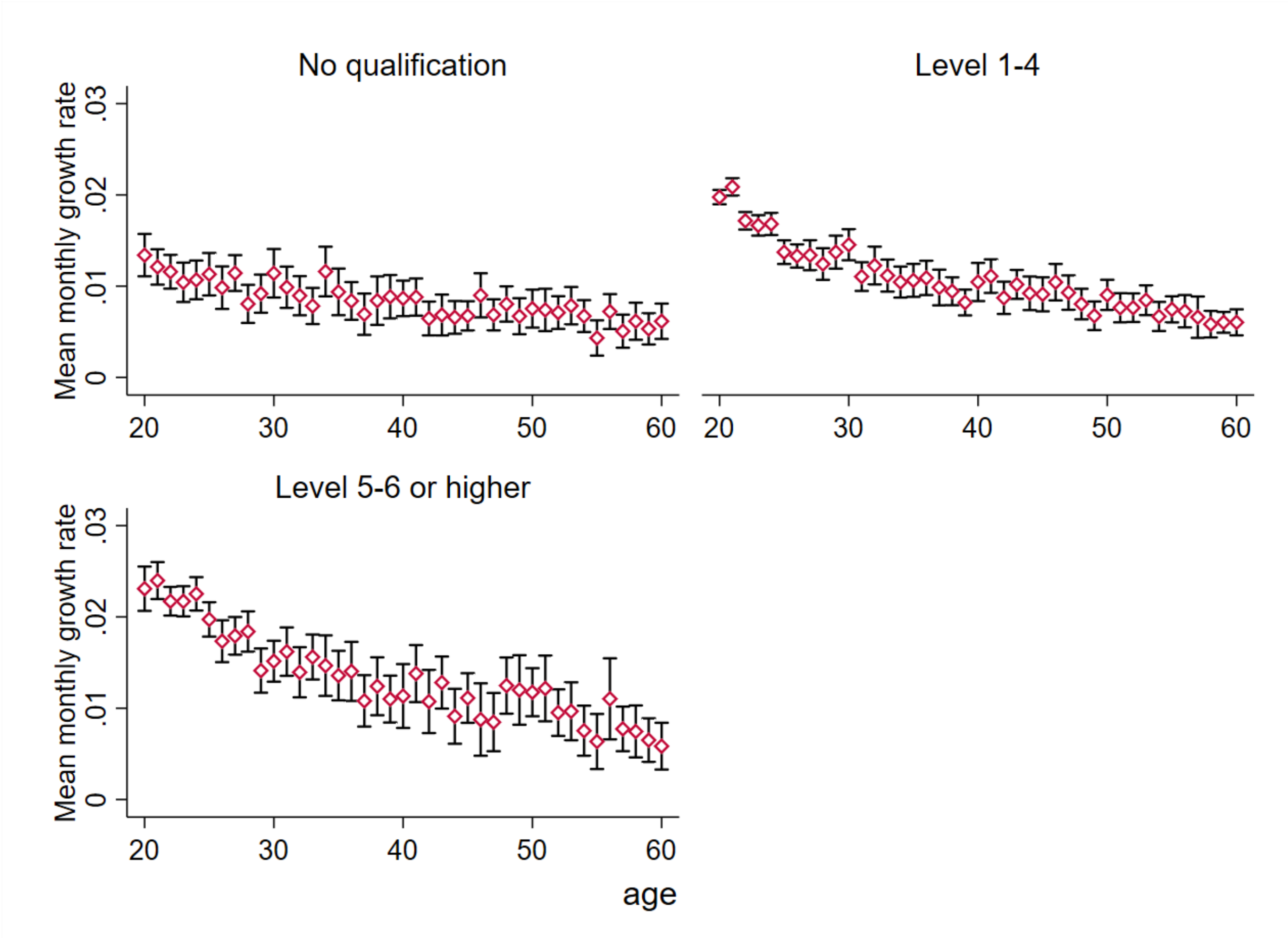



*Note:* The graph shows the accumulated qualification distribution for each age (in years) as at March 2013 for the sample of 26 487 men who were on low pay between January and March 2013.

ure 1 shows the distribution of qualification levels across age. First, the share of workers with no qualification increases with age: while only about 11% of the 25 year old have no qualification, this share peaks at the age of 49 (35%). Likewise, the share of highly qualified workers steadily declines, for example, starting from 34% at 28 years of age to less than 20% for those 50 and above.

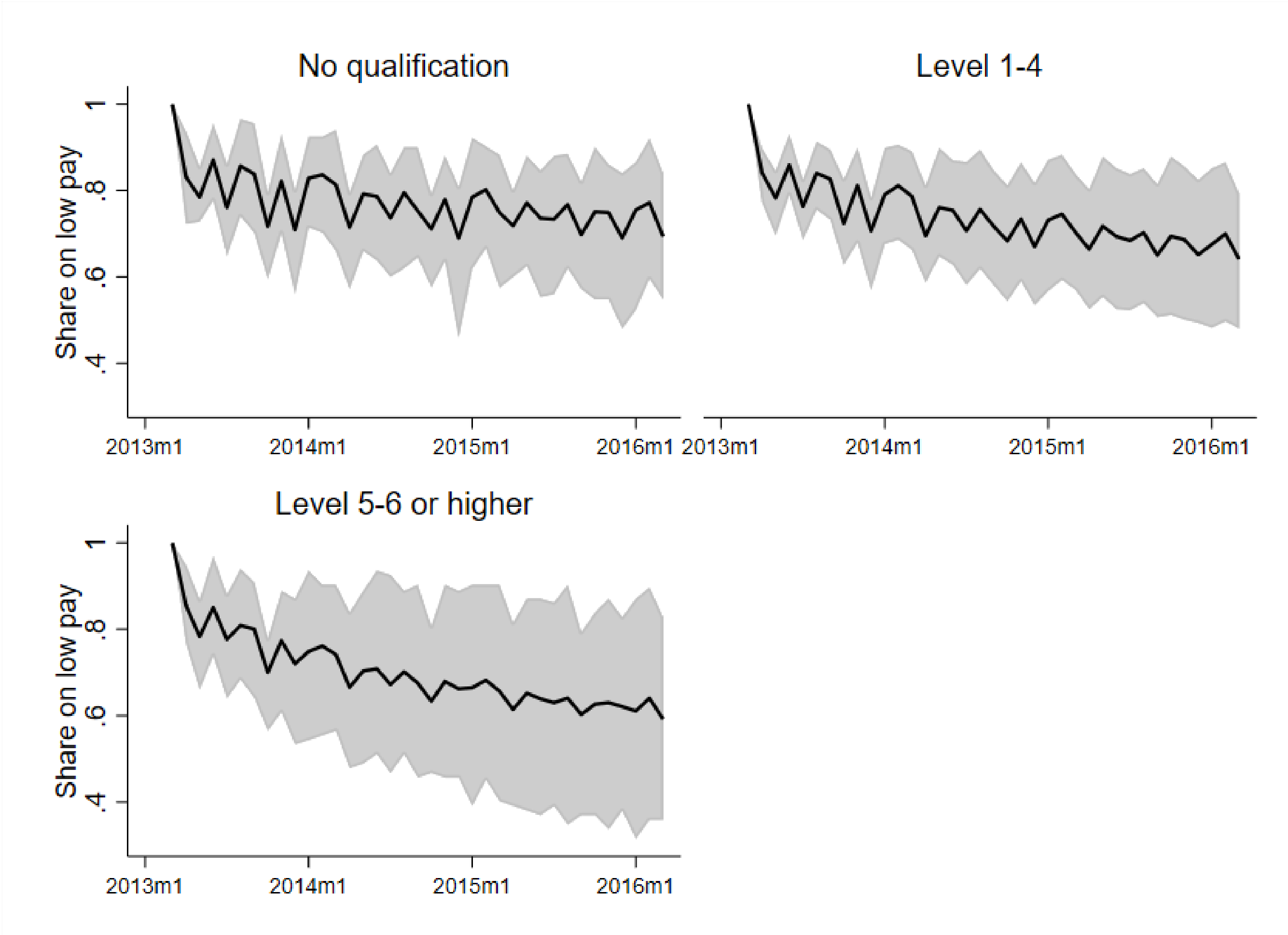
As we have seen from Figure B.1 wage growth is, on average, especially strong for those who are young and highly qualified. For a better understanding regarding the persistence in low-pay we estimate, for each individual, the mean monthly wage growth rate for the period January 2013 to March 2016. For this, we take the monthly difference of each individual's log-wages and, using a simple OLS regression model, we estimate the time trend for each individual. Figure 2 shows the respective age- and qualification-specific mean monthly wage growth rate and the corresponding 95% confidence interval. As Figure 2 highlights, there is substantial heterogeneity not only between qualification levels but as well as within each qualification level across age. For example, the mean monthly wage growth rate for highly qualified workers below 30 is around 2% and drops to 1% (or lower) for those aged 50 and above. However, focusing on to those without any qualification, the heterogeneity in wage growth rate across age is much smaller. On average, younger workers do not experience a substantially greater wage growth rate compared to their older counterparts.

**Figure 2:** Mean monthly income growth rate




*Note:* The graph shows the mean monthly wage growth rate and the respective 95% confidence interval for the sample of 26487 men who were on low pay between January and March 2013, differentiated according to qualification level.

**Figure 3:** Low pay over time




*Note:* The black line shows for the period March 2013 until March 2016 the qualification-group specific mean share of low-paid workers; the shaded area gives the minimum, resp. maximum age-group specific share in the respective month.

Next, we look at the proportion of workers who manage to exit low-pay employment. It is important to note that as we only consider two labour market positions (low pay and higher pay), an individual who exists in low pay is necessarily moving into higher pay. First, we calculate, for each qualification category, the monthly share of workers that are still on low pay. In Figure 3, for each of the three qualification categories, the black line refers to the mean share of individuals who are transitioning into higher-paid jobs. Independent of educational background, we see that over the period of three years this share is constantly declining. For example, in March 2016 about 70% of workers with no qualification still receive low pay whereas the respective numbers are 64% and 59% for the medium qualified (Level 1-4) and highly qualified (Level 5-6 or higher). Moreover, we calculate for each qualification and age combination the respective share at each time point. The shaded area in Figure 3 represents the range of the values of the estimated proportions and portrays that the spread widens over time, especially for the highly qualified. For example, among the highly qualified in March 2016, the age-group specific share of low-pay employment ranges between 36% and 83%. For the group without any qualification, however, the spread ranges from 55% to 84%. The qualification-group specific correlation coefficient between age and share of low pay is strongly positive and increases by qualification level (no qualification: 0.5601; Level 1-4: 0.5999; Level 5-6 and higher: 0.6468).

# 4 Empirical identification strategy

Our population spine consists of New Zealand men who were continuously employed in the period January 2013 to March 2016 and who were on low pay from January until March 2013. To estimate labour market transitions, we follow the economic literature by utilizing dynamic non-linear models (e.g., Stewart, 2007; Buddelmeyer et al., 2010; Clark and Kanellopoulos, 2013). The underlying idea is that the labour market dynamics follow a Markov process of first (or higher) order, which means that the status in the previous period(s) has a *genuine* effect on the position in the subsequent period. Moreover, if individual effects are persistent over time, not accounting for unobserved heterogeneity will lead to an over-estimation of state dependence (Heckman, 1981a; Stewart, 2007).

To start with, we define our variable of interest $y$ such that it takes the value 1 if the individual is on low-pay and 0 otherwise. To estimate state dependence in low pay, the basic dynamic reduced-form model takes the following form:

$$y_{it} = \mathbf{1}\left(\alpha y_{it-1} + \sum_{r=21}^{60} \delta_r(\text{age} = r) + \sum_{s=2}^{3} \gamma_s(\text{qual} = s) + x_i'\beta + \nu_i + u_{it} > 0\right) \quad (1)$$

where the subscripts $i = 1, \ldots, N$ are individuals and $t = 2, \ldots, 36$ is a time identifier on the monthly basis, where $t = 2$ refers to April 2013 and running up to 36 for March 2016. The variable $y_{it-1}$ indicates whether the individual was on low pay in the previous month and thus, $\alpha$ captures the degree of state-dependence in low-pay. Furthermore, $x_i'$ is a vector of individual- and labour market-related explanatory variables. We include the following indicators that are retrieved from the March 2013 Census: ethnicity (categorical: 1 NZ European, 2 Maori, 3 Pacifica, 4 Asian, 5 MELAA, 6 other), smoking regular (dummy), legal marital status (categorical: 1 Married or civil union (not seperated), 2 Seperated, Divorced or dissolved, Widowed or surviving civil union member, 3 Never married and never in a civil union), urban/rural code (categorical: 1 Main Urban Area, 2 Secondary Urban Area, 3 Minor Urban Area 4, Rural Centre 5 Other Rural), North-South Island indicator (dummy). We also include a month indicator and two variables related to the individual's labour market attachment in 2012: one capturing the number of months receiving income from wages and salaries (continuous), another the number of months in low pay (continuous), and an interaction term including both.[6] We also account for age, where $\delta_r$ with $r \in \{21, \ldots, 60\}$ refers to the age-related differences to be on low pay (with age 20 as reference category). Furthermore, we control for qualification-related differences $(\gamma_s)$, where the qualification variable *qual* takes the value 1 if the individual has no qualifications, 2 for Levels 1-4, and 3 for Levels 5-6 and higher (thus, no qualification is the reference category). Note that both variables refer to what was observed during the Census in March 2013, thus age and qualification level are held constant over the observed period. Finally, $\nu_i$ is an individual-specific time-invariant shock[7] and

[6] A standard approach in the economic literature is to include time-varying covariates and to add their time-means (Mundlak, 1978; Chamberlain, 1984). This is not possible in our study as we only include variables that refer to the 2013 Census and labour market performance in 2012. However, we expect that some of the unobserved heterogeneity is captured by running age- and qualification-specific regressions.

[7] A limitation of the model is the auxiliary distribution assumption on the distribution of the random-effects error term. Stewart (2007) has tested the robustness of his findings by applying a dynamic linear probability (DLP) model by using a Arellano and Bond (1991) GMM estimator. One conclusion is that the average partial effects of the lagged labour market positions (in his

$u_{it}$ is an idiosyncratic shock. If individual effects are persistent over time, they are likely to be correlated with the labour market position in the initial period. In the economic literature, this aspect is discussed as the 'initial conditions problem' (Heckman, 1981b; Wooldridge, 2005). However, as we have trimmed the sample to initially low-paid workers, we do not have to further control for this aspect (though we also include labour market variables that refer to 2012).

Here it is worth briefly discussing why trimming the data-set to include only the initially low-paid employed is reasonable. We know from the descriptive statistics that the likelihood of being on low pay is asymmetrically skewed towards younger workers, who also face higher wage growth rates. Thus, we expect that the effect of the lagged labour market position $(\alpha)$ is not independent of the age. However, $\delta_r$ only captures the age related differences of being on low pay at $t$. We would expect that the likelihood of being on low pay is high for younger workers but state dependence is small due to a high wage growth rate. We expect the opposite for older worker.

A limitation of this model is that it only accounts for the labour market status in the previous month $t-1$, although the worker might have been on low pay for a single or multiple month(s). An option to circumvent this issue is to add more lags. However, as we interact the lagged dependent variables, the number of permutations based on past labour market positions increases strongly with the number of lags. We included three lags, which leaves us with eight labor market

study: low pay and unemployment) are higher than in the case of the random-effects models. Own simulation also pointed at the same direction.

**Table 2:** Combinations of lagged labour market position

| | on low pay in | | |
|---|---|---|---|
| $j$ | $t-1$ | $t-2$ | $t-3$ |
| 1 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 |
| 3 | 0 | 1 | 0 |
| 4 | 0 | 1 | 1 |
| 5 | 1 | 0 | 0 |
| 6 | 1 | 0 | 1 |
| 7 | 1 | 1 | 0 |
| 8 | 1 | 1 | 1 |

combinations (see Table 2). To simplify notation, $y_{it-3}$ refers to the categorical variable that holds all combinations of the lagged labour market positions from $t-1$ until $t-3$. The adjusted reduced-form model takes the following form:

$$y_{it} = \mathbf{1}\Bigg( \sum_{j=2}^{8} \alpha_j (y_{it-3} = j) + \sum_{r=21}^{60} \delta_r (\text{age} = r) + \sum_{s=2}^{3} \gamma_s (\text{qual} = s) + x_i' \beta + \nu_i + u_{it} > 0 \Bigg) \quad (2)$$

Reference category is $j = 1$ which refers to not being on low pay in any of the three previous months. Note that by sample design, all individuals are on low pay from January to March 2013 and thus start with $j = 8$. In the current specification, it is assumed that the effect of the lagged labour market positions $(y_{it-3})$ is independent of time $t$. However, as shown in Figure 3, there is descriptive evidence that with the elapsed time the share of low-paid workers declines on average. We extend Equation 2 accordingly by adding a time trend $\lambda_t$:

$$y_{it} = \mathbf{1}\Bigg(\sum_{j=2}^{8} \alpha_j (y_{it-3} = j) + \lambda_t + \sum_{r=21}^{60} \delta_r (\text{age} = r) + \sum_{s=2}^{3} \gamma_s (\text{qual} = s) + x_i'\beta + \nu_i + u_{it} > 0\Bigg) \quad (3)$$

So far, Equation 3 assumes that the effect of the time trend $\lambda_t$ is independent of past labour market status. However, the chances of exiting low pay may decline with time spent in the low-pay sector, so we extend the model by adding an interaction term between the time trend and the lagged labour market position. Our model takes the following form:

$$y_{it} = \mathbf{1}\Bigg(\sum_{j=2}^{8} \alpha_j (y_{it-3} = j) + \lambda_t + \sum_{j=2}^{8} \theta_j (y_{it-3} = j)\lambda_t + \sum_{r=21}^{60} \delta_r (\text{age} = r) + \sum_{s=2}^{3} \gamma_s (\text{qual} = s) + x_i'\beta + \nu_i + u_{it} > 0\Bigg) \quad (4)$$

Because being continuously on higher pay ($j = 1$) is the reference category, $\theta_j$ represents the evolutionary effect of lagged labour market positions over time. Thus, the probability of staying on low pay at time point $t$ is the combined effect of an underlying effect $\alpha_j$ and its time deviation $\lambda_t + \theta_j \lambda_t$. Based on our descriptive findings in Figure 3, we expect that low pay persistence is heterogeneous across the workforce and should decline more intensely for young and qualified workers. This means, we expect that $\alpha_j$ and $|\theta_j|$ increase with age and being less qualified. To account for age- and qualification-related differences in low pay state-dependence, one option is to include respective interaction terms. However, one downside of this approach is that this implicitly assumes that the effect of the covariates and the individual-specific effect $\nu_i$ are independent of age and qualification level. A less restrictive approach is to run age-qualification specific models. The final specification takes the following form:

$$y_{it}^{a,q} = \mathbf{1}\left(\sum_{j=2}^{8} \alpha_j (y_{it-3} = j) + \lambda_t + \sum_{j=2}^{8} \theta_j (y_{it-3} = j)\lambda_t + x_i'\beta + \nu_i + u_{it} > 0\right) \quad (5)$$

with $y_{it}^{a,q}$ representing the low-paid indicator of an individual $i$ at month $t$ for a unique combination of age $a$ and qualification $q$. Our sample consists of a large number of individuals but only a small number of time-points, therefore asymptotics are on $N^{a,q}$, the number of age-qualification specific observations, alone. We assume that both error terms follow a normal distribution, e.g., $\nu_i \sim N(0, \sigma_\nu^2)$ and $u_{it} \sim N(0, \sigma_u^2)$ and that $u_{it}$ is iid. As the outcome variable $y_{it}^{a,q}$ is dichotomous, a normalization of $u_{it}$ is required. We chose $u_{it} \sim N(0,1)$ and the age-qualification specific outcome probability is:

$$\begin{aligned} P_{it}^{a,q}(\nu^*) = \Phi\Bigg[(2y_{it} - 1)\Bigg(\sum_{j=2}^{8} \alpha_j (y_{it-3} = j) + \lambda_t \\ + \sum_{j=2}^{8} \theta_j (y_{it-3} = j)\lambda_t + x_i'\beta + \sigma_\nu \nu^*\Bigg)\Bigg] \end{aligned} \quad (6)$$

Note that $\Phi[\cdot]$ refers to the cumulative standard normal distribution. The age-qualification specific likelihood function is the product of all time-point specific probabilities across all individuals of same age and qualification. Namely,

$$L^{a,q} = \prod_{i=1}^{N^{a,q}} \int_{v^*} \left\{ \prod_{t=2}^{36} P_{it}^{a,q}(v^*) \right\} \mathrm{d}F(v^*) \tag{7}$$

where $F$ is the distribution function of $v^* = {}^{v}/_{\sigma_v}$. Equation (7) does not have a closed-form expression, and therefore $v$ has to be integrated out. As we assume that $v$ is normally distributed, the integral can be evaluated using Gaussian-Hermite quadrature (Butler and Moffitt, 1982).

In total, we estimate 123 regressions as we have 41 age bins and 3 qualification groups. To make the findings comparable, we estimate the partial effect of staying on low pay for each combination. We look at how the risk of staying on low pay changes after 12 months given that the individual was on low pay in each of the three previous months:

$$\begin{aligned} PE_i^{a,q} = \Phi\left[\left(\hat{\alpha}_8 + \hat{\lambda}_{14} + \hat{\theta}_8\hat{\lambda}_{14} + x_i'\hat{\beta}\right)\left(\sqrt{1-\hat{\lambda}}\right)\right] \\ - \Phi\left[\left(\hat{\alpha}_8 + \hat{\lambda}_2 + \hat{\theta}_8\hat{\lambda}_2 + x_i'\hat{\beta}\right)\left(\sqrt{1-\hat{\lambda}}\right)\right] \end{aligned} \tag{8}$$

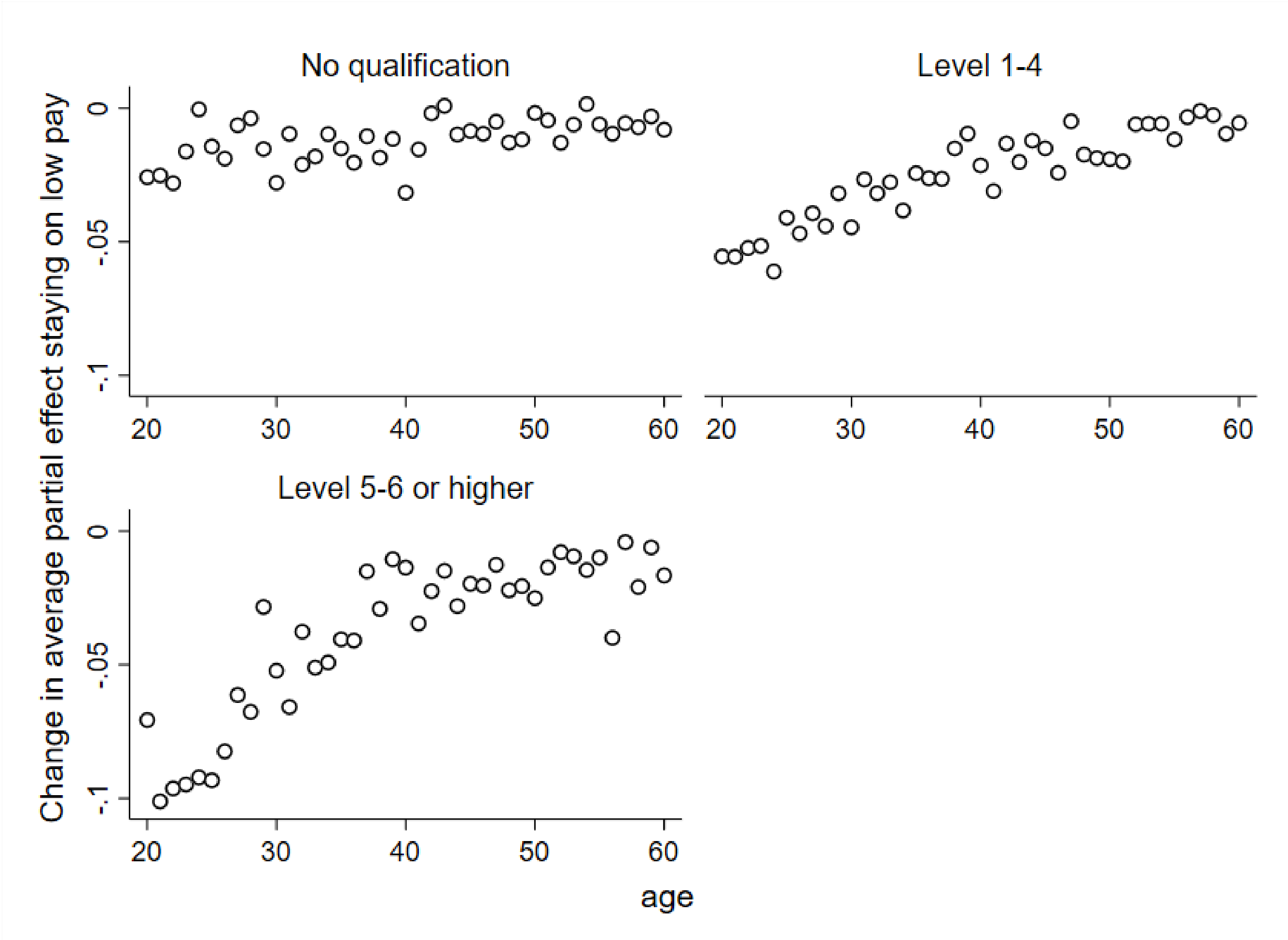
with $\hat{\lambda} = {}^{\hat{\sigma}_v}/_{(\hat{\sigma}_v+1)}$ (Arulampalam, 1999).

# 5 Results

In this study, we analyse how changes in low-pay persistence varies with age and qualification. To make the heterogeneous effects across different age bins and qualification levels visible, we calculate for each combination partial effects of how the risk of staying on low pay changes with elapsed time. Descriptive evidence points at a higher mobility for younger and qualified workers compared to non-qualification workers.

We start with calculating the probability of being on low pay when an individual has worked in the low-pay sector continuously for the last three months (see Figure B.4). We calculate the probabilities for April 2013 ($t = 2$) and twelve months later for April 2014 ($t = 14$). We see that, on average, the probability of staying on low pay ranges between 70 and >90%, with higher probabilities for older workers and for less qualified workers. Moreover, we can see that the probability of staying on low pay is lower 12 months later, though this difference varies across the different age and qualification combinations. For example, the gap is clearly visible for young and (highly) qualified low-paid workers, but for older workers without any qualification the probabilities for the two time points are almost identical.

**Figure 4:** Average partial effect




*Note:* The graph shows for each age (in years as at March 2013) and differentiated by qualification group, the mean change between April 2014 and April 2013 in the probability staying on low pay when working on low pay in the three previous months ($j = 8$). For example, the probability staying on low pay declines by, on average, ten percentage points between April 2014 and April 2013 for someone who was low-paid employed in the three previous months and is 21 years old with a high qualification (Level 5-6 or higher as at March 2013).

Next, we calculate the difference between the two different time points. Figure 4 shows for the three qualification groups how, on average, the risk of staying on low pay changes after twelve months for an individual who was working in the low-pay sector in the previous three months. The mean values indicate large degrees of heterogeneity, both at the age level as well as between qualifications. For example, for an individual in his early twenties and holding a Level 5-6 or higher qualification the risk of staying on low pay declines, on average, by eight to ten percentage points. This number drops to, on average, one to two percentage points for men above 50. When we move down the qualification ladder, we can see that the difference across the different age bins is substantially lower than for those without any qualifications. Going back to the two specific age groups, the respective numbers for the group in their early twenties are between two and three percentage points and for those over 50 are around one percentage point.

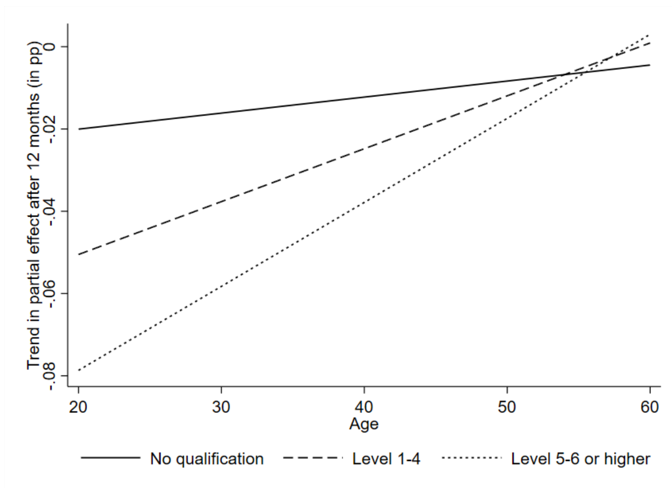
We also calculated the slope of the average partial effects across age and qualification group. We apply a simple OLS model where our dependent variable is the age and qualification specific average partial effect. As explanatory variables, we include age as a continuous variable, the qualification level as a categorical variable and an interaction of both terms (see Model (1) in Table 3). To visualize the relationship, we predict the qualification specific slopes across age. Figure 5 shows a positive slope for all qualification levels, which means that the decline in low-pay persistence within 12 months is stronger for younger workers than older workers. However, the slope is much steeper for those with high qualification levels, with a magnitude up to four times higher (-0.02pp for a 20 year old worker with no qualification vs. -0.08pp for a 20 year old worker holding a Level 5-6 or higher qualification).

To highlight the effect of age-qualification specific effects, we re-run our estimation in a pooled sample where we account for age and qualification level as covariates. Figure B.3 shows the corresponding predicted slopes across the average partial effects. We find that these slopes are also positive, however the effect is much weaker compared to the case we run age-qualification specific regressions. Furthermore, we can see small changes across different qualification levels, however the magnitude of these is very small as well. We do not observe a single horizontal line as we have not interacted age and qualification with the lagged labour market position and the time trend. It needs to be kept in mind that older and less qualified workers have a higher likelihood of staying on low pay and thus the marginal effect of the time trend decreases.[8]

We have presented empirical evidence that, with time, the risk of staying on low pay drops more sharply for younger and higher qualified workers. This finding is consistent with the observation of the wage pattern over the life-cycle, which shows that wage growth is, on average, especially strong for younger and higher qualified workers. The economic literature points at, among other factors, two determinants for wage growth: on-the-job training and improved employer-employee matches. In the following, we want to test whether we find indications for improved employer-employee matches. As a proxy, we take advantage of the tax records, which also holds an employer identifier. For each low-paid worker,

[8]For example, $\Phi(x_1+a)-\Phi(x_1)<\Phi(x_2+a)-\Phi(x_2)$ if $x_1>x_2>0$ and $a>0$.

**Figure 5:** Trend in average partial effects




*Note:* The graph shows the trend in the average partial effects between April 2014 and April 2013 for workers staying on low pay after working three months on low pay. For example, for a 20 year old (as at March 2013) highly qualified worker, the probability of staying on low pay when they have worked for three months on low pay declines, on average, by 8% points after one year.

we calculate the mean wage the respective employer has paid to all of his staff. If a low-paid worker holds multiple jobs, we only account for the job with the highest wage. For confidentiality reasons, we only include employers with a minimum of five employees. Next, we calculate the mean wage of the employer for each age-qualification combination.

In Figure B.5, the distribution of the log mean wages are shown for March 2013, the start of observation period, and March 2016, the last month of our observation period (numbers are in nominal terms as we are not interested in quantifying the changes). When looking at the mean wage distribution for March 2013, we do not observe any specific trend across age and qualification level. However, when moving to the March 2016 distribution, we can see that, independent of the qualification level, there is a negative association by age, indicating that younger workers tend to work at higher paying firms. Moreover, the slope is more pronounced for higher qualification levels.

We also run an OLS regression for both time points, where the mean wage of the employer for each age-qualification combination is the dependent variable and age and qualification level (and their interaction) as independent variables. The first column of Table 4 shows the respective coefficients for March 2013. Though we can see a negative slope across age, the magnitude is small (a drop of 0.16pp in 10 years) and not significantly different from zero. Furthermore, the regression model indicates that those with a higher qualification level work for an employer who pays a somewhat lower mean wage, but again the magnitude is small (-0.4pp for a Level 1-4 qualification and -2.5pp for a Level 5-6 and higher qualification) and statistically insignificant. The only significant effect (despite the constant) is a qualification-specific age effect, indicating that the mean wage of the employer increases for the highest qualification group over time (an additional 2.3pp in 10 years).

The model provides substantially different results when looking at March 2016, both in magnitude and direction. First, we still find a negative association with age, though the magnitude is much stronger (-3pp in 10 years) and significantly different from zero. Furthermore, we see that workers who are highly qualified work in firms with a substantially higher mean wage (+10.9pp for Levels 1-4, and +25.3pp for Levels 5-6 or higher). Lastly, the slope is steeper for highly qualified groups which shows that young and highly qualified workers manage to move into higher paying firms more so than other groups.

To complement the picture on low-pay transition, we also looked at the share of workers still working in the low-pay sector five years later in the Census 2018. We calculate the wage distribution for 20 to 60 year old male workers in March 2018 and define the two lowest deciles as low-pay workers. Next, for each age-qualification combination, we calculate the share of workers who are still working in the low-pay sector five years later (we exclude those without any income from wages and salaries in 2018). Like before, to analyse the relationship, we apply a simple OLS model where the share on low pay is the dependent variable and age and qualification level (and their interaction) are the covariates. To ensure consistency, we only account for age range of 20 to 55 years (inclusive). The third column of Table 3 presents the respective coefficients. We find that with each year of life, the share of initially low-paid workers that are still working in the low-pay sector increases by 0.9pp. Moreover, we find that qualified workers are noticeably less likely to remain on low pay (-21pp for Levels 1-4 and -38.8pp for Levels 5-6 and higher). Finally, we also find that the age effect is more pronounced for higher qualification levels. For example, for those with Level a 5-6 or higher qualification, each additional year of age increases the low-pay share by additional 0.6pp.

# 6 Conclusion

To estimate how low-pay persistence changes over time for different cohorts and levels of human capital, we run age- and qualification-specific regressions and control for time trends in state dependence. Moreover, in contrast to previous studies which looked at annual labour market transitions, we employ administrative tax records to track monthly earnings. Our findings show that persistence in low-pay drops the most for young qualified workers, whereas low-pay persistence is almost constant over time and without variation for older workers regardless of the qualification levels. This finding is in line with the literature on human capital formation, which explains why wages, on average, take a hump-shaped pattern over the life-cycle. Further, we find evidence that young and qualified workers, on average, manage to transit into higher-paying firms more frequently than do their less qualified or older colleagues. We interpret this finding as an indication of improved employer-employee matches.

These findings highlight the importance of controlling for heterogeneity across the workforce. With respect to the likelihood of exiting low-pay employment, accounting for differences in age and qualification by including them as covariates is not sufficient. Likewise, caution needs to be exercised when making policy recommendations. As the heterogeneity in the earnings prospects revealed, there cannot be a 'one-size-fits-all' policy. Being young does not prevent an individual from being stuck in the low-pay sector, as shown for those young workers with no qualification. However, having a high qualification is also not a sufficient protection against being trapped in low pay as shown for the group of workers above 50. However, being on low pay does not mean to be deemed staying on the wage level for the subsequent periods.

A limitation of this study is that it only considers transitions between low and higher pay. A large body of literature compares the prospects of those on low pay with those who are unemployed. The underlying question is whether low-pay employment acts as a springboard into higher-pay employment and offers greater opportunity to climb up the earnings ladder compared to remaining unemployed. Like low-pay employment, it can also be argued that employment and earning prospects of unemployed differ across age and qualification level(s). Another useful extension might be to decompose higher-paid group further. So far, we lump all individuals above the low-pay threshold together. However, due to the differences in wage growth, young and more qualified workers might be more likely to enter higher parts of the earnings distribution and older workers might be stuck just above the cut-off point.

# References

J. M. Abowd, F. Kramarz, and D. N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.

R. Alessie, S. Hochguertel, and A. v. Soest. Ownership of stocks and mutual funds: A panel data analysis. *Review of Economics and Statistics*, 86(3):783–796, 2004.

M. Arellano and S. Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297, 1991.

W. Arulampalam. A note on estimated coefficients in random effects probit models. *Oxford Bulletin of Economics and Statistics*, 61(4):597–602, 1999.

W. Arulampalam, A. L. Booth, and M. P. Taylor. Unemployment persistence. *Oxford economic papers*, 52(1):24–50, 2000.

S. Ayllón. Youth poverty, employment, and leaving the parental home in e urope. *Review of Income and Wealth*, 61(4):651–676, 2015.

M. Bhuller, C. N. Brinch, and S. Königs. Time aggregation and state dependence in welfare receipt. *The Economic Journal*, 127(604):1833–1873, 2017.

M. Biewen. Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. *Journal of Applied Econometrics*, 24(7):1095–1116, 2009.

J. N. Brown. Why do wages increase with tenure? on-the-job training and life-cycle wage growth observed within firms. *The American Economic Review*, pages 971–991, 1989.

H. Buddelmeyer, W.-S. Lee, and M. Wooden. Low-paid employment and unemployment dynamics in australia. *Economic Record*, 86(272):28–48, 2010.

J. S. Butler and R. Moffitt. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50(3):761–764, 1982.

L. Cai. Is there a low-pay no-pay cycle in australia? a note on fok, scutella and wilkins (2015). *Empirical Economics*, pages 1–19, 2019.

L. Cai, K. Mavromaras, and P. Sloane. Low paid employment in britain: Estimating state-dependence and stepping stone effects. *Oxford Bulletin of Economics and Statistics*, 80(2):283–326, 2018.

L. Cappellari. Earnings mobility among italian low-paid workers. *Journal of Population Economics*, 20(2):465–482, 2007.

G. Chamberlain. Panel data. In Z. Griliches and M. Intriligator, editors, *Handbook of econometrics*, pages 1247–1318. North Holland, 1984.

K. Clark and N. C. Kanellopoulos. Low pay persistence in europe. *Labour Economics*, 23:122–134, 2013.

S. Dimitrakopoulos and M. Kolossiatis. State dependence and stickiness of sovereign credit ratings: evidence from a panel of countries. *Journal of Applied Econometrics*, 31(6):1065–1082, 2016.

European Commission. Comparative analysis of the european qualifications framework and the new zealand qualifications framework: joint technical report, 2017.

H. S. Farber and R. Gibbons. Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047, 1996.

Y. K. Fok, R. Scutella, and R. Wilkins. The low-pay no-pay cycle: Are there systematic differences across demographic groups? *Oxford Bulletin of Economics and Statistics*, 77(6):872–896, 2015.

P. Haan and M. Myck. Dynamics of health and labor market risks. *Journal of health economics*, 28(6):1116–1125, 2009.

J. J. Heckman. Heterogeneity and state dependence. In S. Rosen, editor, *Studies in labor markets*, pages 91–140. University of Chicago Press, 1981a.

J. J. Heckman. The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process. In C. F. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 179–195. MIT Press, 1981b.

D. Lagakos, B. Moll, T. Porzio, N. Qian, and T. Schoellman. Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849, 2018.

H. Low, C. Meghir, and L. Pistaferri. Wage risk and employment risk over the life cycle. *American Economic Review*, 100(4):1432–67, 2010.

B. McCormick. A theory of signalling during job search, employment efficiency, and "stigmatised" jobs. *The Review of Economic Studies*, 57(2):299–313, 1990.

R. Mendes, G. J. Van Den Berg, and M. Lindeboom. An empirical assessment of assortative matching in the labor market. *Labour Economics*, 17(6):919–929, 2010.

D. Metcalf. The low pay commission and the national minimum wage. *The Economic Journal*, 109(453):46–66, 1999.

A. Mosthaf. Do scarring effects of low-wage employment and non-employment differ between levels of qualification? *Scottish Journal of Political Economy*, 61(2):154–177, 2014.

Y. Mundlak. On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85, 1978.

A. Plum. The british low-wage sector and the employment prospects of the unemployed. *Applied Economics*, 51(13):1411–1432, 2019.

A. Skrondal and S. Rabe-Hesketh. Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pages 211–237, 2014.

P. J. Sloane and I. Theodossiou. Earnings mobility, family income and low pay. *The Economic Journal*, 106(436):657–666, 1996.

M. B. Stewart. The interrelated dynamics of unemployment and low-wage employment. *Journal of Applied Econometrics*, 22(3):511–531, 2007.

M. B. Stewart and J. K. Swaffield. Low pay dynamics and transition probabilities. *Economica*, 66(261):23–42, 1999.

R. H. Topel and M. P. Ward. Job mobility and the careers of young men. *The Quarterly Journal of Economics*, 107(2):439–479, 1992.

J. M. Wooldridge. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1):39–54, 2005.

C. Wunder and R. T. Riphahn. The dynamics of welfare entry and exit amongst natives and immigrants. *Oxford Economic Papers*, 66(2):580–604, 2014.

# A Disclaimer

The results in this paper are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics NZ.

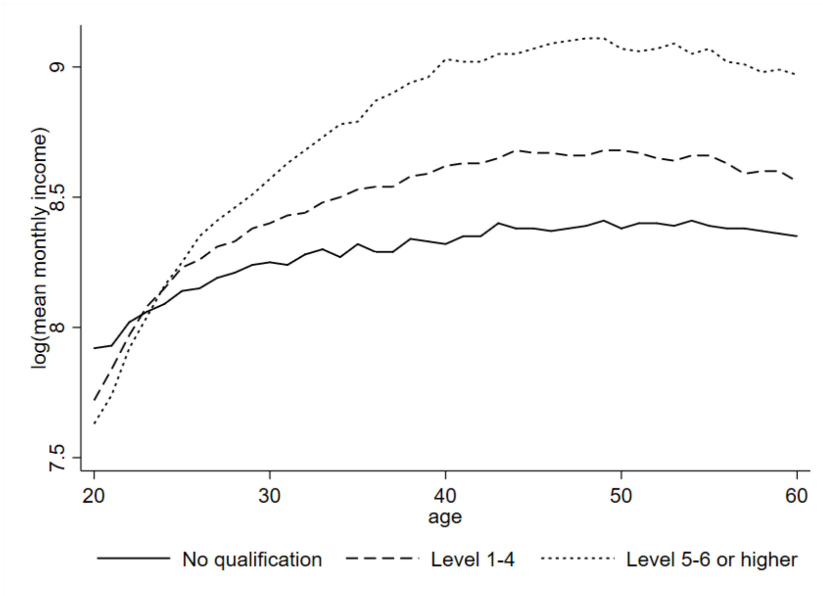
The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes. Any person who has had access to the unit record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification. Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI.

Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

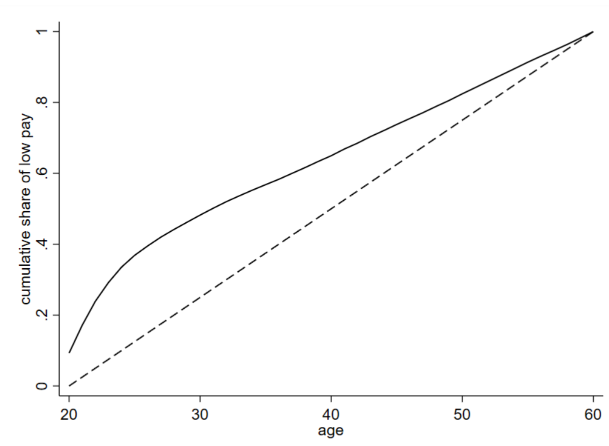
# B Figures

**Figure B.1:** Income distribution



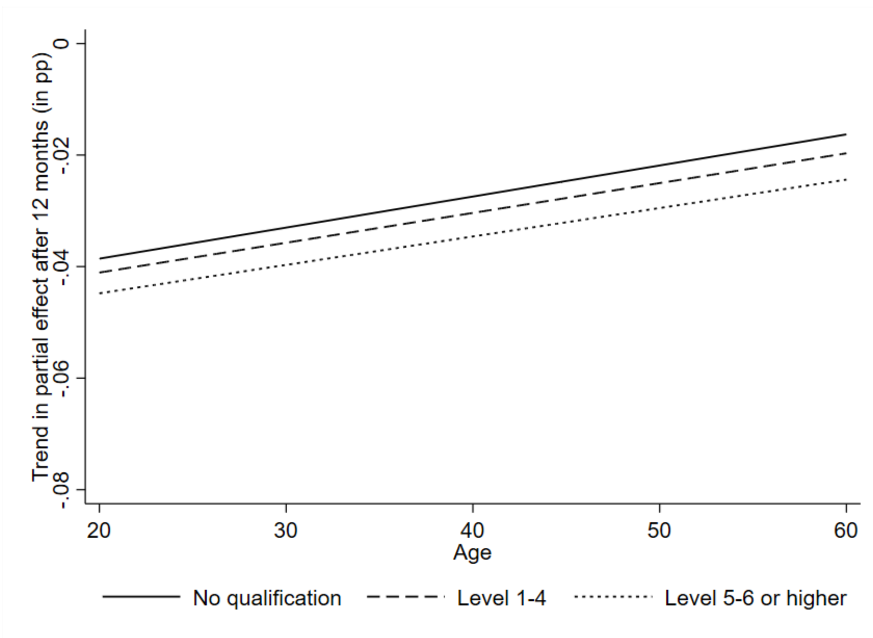

*Notes:* The graph shows for March 2013 the distribution of log-income from wages & salaries of 601 686 men aged 20 to 60 who received income in each month of the period January 2013 to March 2016.

**Figure B.2:** Age distribution of low-pay worker.



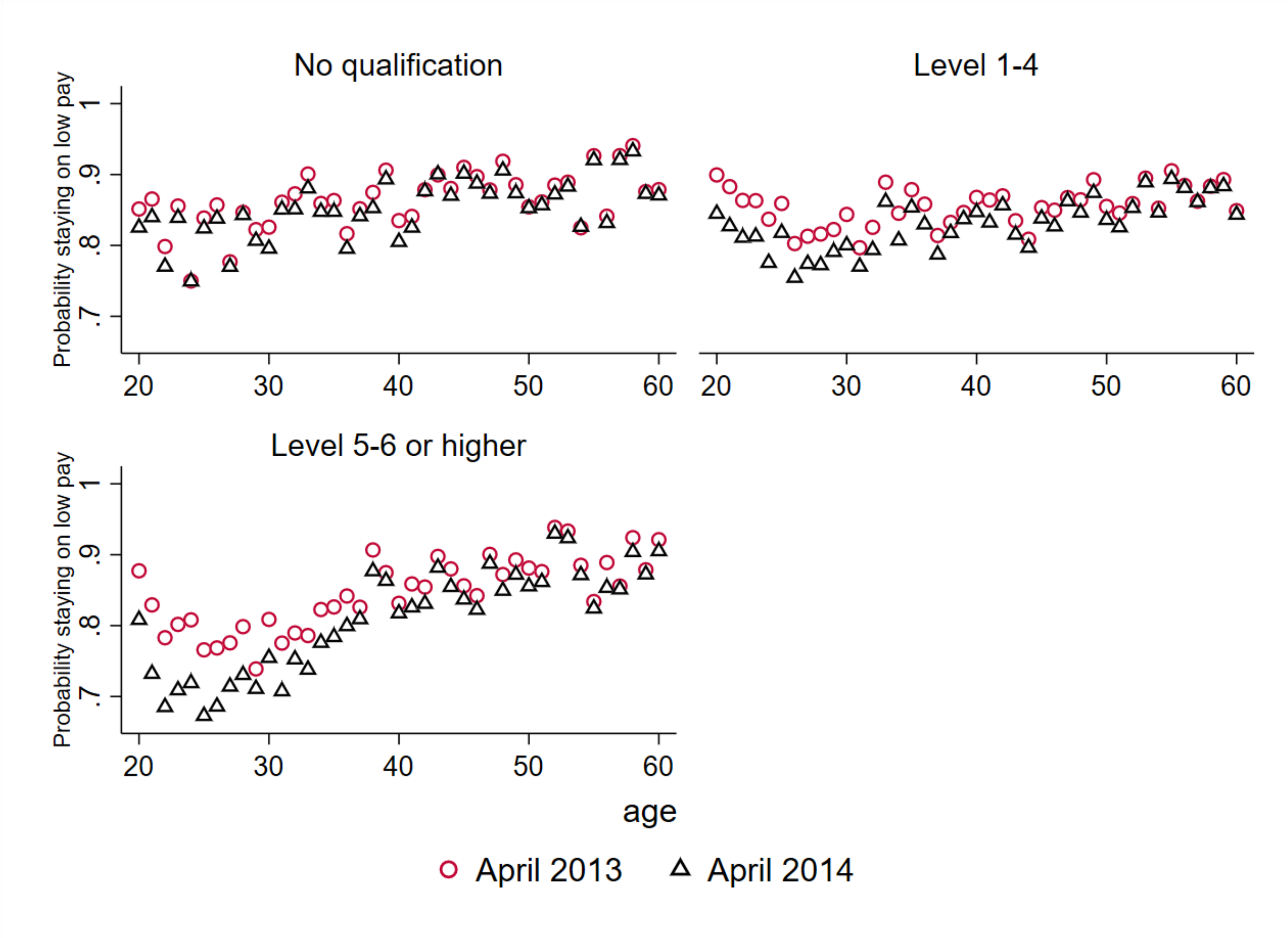

*Note:* The black line shows the accumulated age distribution for the sample of 26487 men who were on low pay between January and March 2013. The dashed line indicates equal distribution across age.

**Figure B.3:** Trend in average partial effects (pooled sample)




*Note:* The graph shows the trend in the average partial effects between April 2014 and April 2013 for staying on low pay after working three months on low pay when the regression is based on a pooled sample. For example, for a 20 year (as at March 2013) old highly qualified worker the probability to stay on low pay when having worked for three months on low pay declines, on average, by 4.2% points after one year.

**Figure B.4:** Probability of staying on low pay




*Note:* The graph shows the mean probability for each age (in years as at March 2013), differentiated by qualification group for April 2013 (red circles) and April 2014 (black triangles).

**Figure B.5:** Wage distribution of employer


*Note:* The graph shows for each age (in years as at March 2013) and differentiated by qualification group, the mean log wages (nominal) of the employer for March 2013 (red circles) and March 2016 (black diamonds).

# C Tables

**Table 1:** Distribution of income from wages and salaries

| | No qualification | Level 1-4 | Level 5-6 or higher | | No qualification | Level 1-4 | Level 5-6 or higher |
|---|---|---|---|---|---|---|---|
| 20 | 7.92 | 7.72 | 7.63 | 41 | 8.35 | 8.63 | 9.02 |
| | (7.35) | (7.23) | (7.19) | | (7.94) | (8.32) | (8.97) |
| 21 | 7.93 | 7.84 | 7.74 | 42 | 8.35 | 8.63 | 9.02 |
| | (7.19) | (7.34) | (7.3) | | (7.77) | (8.37) | (8.87) |
| 22 | 8.02 | 7.97 | 7.92 | 43 | 8.4 | 8.65 | 9.05 |
| | (7.31) | (7.37) | (7.38) | | (8.31) | (8.44) | (9) |
| 23 | 8.06 | 8.08 | 8.04 | 44 | 8.38 | 8.68 | 9.05 |
| | (7.29) | (7.46) | (7.42) | | (7.85) | (8.48) | (8.91) |
| 24 | 8.09 | 8.15 | 8.16 | 45 | 8.38 | 8.67 | 9.07 |
| | (7.34) | (7.5) | (7.61) | | (7.95) | (8.47) | (9.1) |
| 25 | 8.14 | 8.23 | 8.25 | 46 | 8.37 | 8.67 | 9.09 |
| | (7.43) | (7.53) | (7.59) | | (7.81) | (8.44) | (9.07) |
| 26 | 8.15 | 8.26 | 8.35 | 47 | 8.38 | 8.66 | 9.1 |
| | (7.41) | (7.54) | (7.89) | | (7.7) | (8.42) | (9.14) |
| 27 | 8.19 | 8.31 | 8.41 | 48 | 8.39 | 8.66 | 9.11 |
| | (7.48) | (7.59) | (7.86) | | (7.74) | (8.4) | (9.21) |
| 28 | 8.21 | 8.33 | 8.46 | 49 | 8.41 | 8.68 | 9.11 |
| | (7.47) | (7.65) | (7.85) | | (7.86) | (8.6) | (9.19) |
| 29 | 8.24 | 8.38 | 8.51 | 50 | 8.38 | 8.68 | 9.07 |
| | (7.63) | (7.68) | (8.13) | | (7.89) | (8.57) | (9.07) |
| 30 | 8.25 | 8.4 | 8.57 | 51 | 8.4 | 8.67 | 9.06 |
| | (7.72) | (7.8) | (8.03) | | (8.2) | (8.45) | (8.98) |
| 31 | 8.24 | 8.43 | 8.63 | 52 | 8.4 | 8.65 | 9.07 |
| | (7.48) | (7.86) | (8.35) | | (8.11) | (8.47) | (9.09) |
| 32 | 8.28 | 8.44 | 8.68 | 53 | 8.39 | 8.64 | 9.09 |
| | (7.73) | (7.84) | (8.21) | | (7.96) | (8.45) | (9.24) |
| 33 | 8.3 | 8.48 | 8.73 | 54 | 8.41 | 8.66 | 9.05 |
| | (7.58) | (8.03) | (8.51) | | (7.98) | (8.67) | (9.07) |
| 34 | 8.27 | 8.5 | 8.78 | 55 | 8.39 | 8.66 | 9.07 |
| | (7.55) | (8.09) | (8.48) | | (7.94) | (8.65) | (9.14) |
| 35 | 8.32 | 8.53 | 8.79 | 56 | 8.38 | 8.63 | 9.02 |
| | (7.78) | (8.02) | (8.34) | | (7.84) | (8.46) | (9.23) |
| 36 | 8.29 | 8.54 | 8.87 | 57 | 8.38 | 8.59 | 9.01 |
| | (7.55) | (8.07) | (8.67) | | (8.35) | (8.41) | (8.97) |
| 37 | 8.29 | 8.54 | 8.9 | 58 | 8.37 | 8.6 | 8.98 |
| | (7.65) | (8.01) | (8.68) | | (7.8) | (8.55) | (8.89) |
| 38 | 8.34 | 8.58 | 8.94 | 59 | 8.36 | 8.6 | 8.99 |
| | (7.78) | (8.23) | (8.88) | | (7.88) | (8.51) | (8.99) |
| 39 | 8.33 | 8.59 | 8.96 | 60 | 8.35 | 8.56 | 8.97 |
| | (7.63) | (8.2) | (8.83) | | (7.99) | (8.37) | (8.91) |
| 40 | 8.32 | 8.62 | 9.03 | | | | |
| | (7.64) | (8.41) | (10.01) | | | | |

**Table 2:** Average partial effect of staying low pay

| | No qualification | Level 1-4 | Level 5-6 or higher | | No qualification | Level 1-4 | Level 5-6 or higher |
|---|---|---|---|---|---|---|---|
| 20 | -0.026 | -0.056 | -0.071 | 41 | -0.015 | -0.031 | -0.035 |
| | (0.008) | (0.003) | (0.011) | | (0.007) | (0.006) | (0.012) |
| 21 | -0.025 | -0.056 | -0.101 | 42 | -0.002 | -0.013 | -0.023 |
| | (0.008) | (0.003) | (0.01) | | (0.007) | (0.006) | (0.01) |
| 22 | -0.028 | -0.052 | -0.096 | 43 | 0.001 | -0.02 | -0.015 |
| | (0.01) | (0.004) | (0.009) | | (0.005) | (0.006) | (0.007) |
| 23 | -0.016 | -0.052 | -0.095 | 44 | -0.01 | -0.012 | -0.028 |
| | (0.008) | (0.005) | (0.009) | | (0.006) | (0.007) | (0.011) |
| 24 | 0.000 | -0.061 | -0.092 | 45 | -0.009 | -0.015 | -0.02 |
| | (0.011) | (0.006) | (0.009) | | (0.005) | (0.006) | (0.01) |
| 25 | -0.014 | -0.041 | -0.093 | 46 | -0.01 | -0.024 | -0.02 |
| | (0.01) | (0.005) | (0.011) | | (0.006) | (0.007) | (0.01) |
| 26 | -0.019 | -0.047 | -0.082 | 47 | -0.005 | -0.005 | -0.013 |
| | (0.009) | (0.007) | (0.011) | | (0.006) | (0.005) | (0.008) |
| 27 | -0.006 | -0.039 | -0.061 | 48 | -0.013 | -0.017 | -0.022 |
| | (0.01) | (0.007) | (0.01) | | (0.005) | (0.006) | (0.009) |
| 28 | -0.004 | -0.044 | -0.068 | 49 | -0.012 | -0.019 | -0.021 |
| | (0.009) | (0.008) | (0.011) | | (0.006) | (0.006) | (0.01) |
| 29 | -0.015 | -0.032 | -0.028 | 50 | -0.002 | -0.019 | -0.025 |
| | (0.009) | (0.008) | (0.01) | | (0.007) | (0.006) | (0.009) |
| 30 | -0.028 | -0.045 | -0.052 | 51 | -0.005 | -0.02 | -0.014 |
| | (0.01) | (0.008) | (0.011) | | (0.006) | (0.006) | (0.008) |
| 31 | -0.01 | -0.027 | -0.066 | 52 | -0.013 | -0.006 | -0.008 |
| | (0.008) | (0.007) | (0.013) | | (0.006) | (0.005) | (0.005) |
| 32 | -0.021 | -0.032 | -0.038 | 53 | -0.006 | -0.006 | -0.009 |
| | (0.008) | (0.008) | (0.011) | | (0.005) | (0.004) | (0.005) |
| 33 | -0.018 | -0.028 | -0.051 | 54 | 0.001 | -0.006 | -0.015 |
| | (0.008) | (0.006) | (0.014) | | (0.007) | (0.005) | (0.009) |
| 34 | -0.01 | -0.038 | -0.049 | 55 | -0.006 | -0.012 | -0.01 |
| | (0.008) | (0.008) | (0.013) | | (0.004) | (0.004) | (0.01) |
| 35 | -0.015 | -0.024 | -0.041 | 56 | -0.01 | -0.003 | -0.04 |
| | (0.008) | (0.007) | (0.012) | | (0.007) | (0.004) | (0.015) |
| 36 | -0.02 | -0.026 | -0.041 | 57 | -0.006 | -0.001 | -0.004 |
| | (0.01) | (0.007) | (0.012) | | (0.004) | (0.005) | (0.008) |
| 37 | -0.011 | -0.026 | -0.015 | 58 | -0.007 | -0.003 | -0.021 |
| | (0.008) | (0.007) | (0.01) | | (0.004) | (0.004) | (0.009) |
| 38 | -0.019 | -0.015 | -0.029 | 59 | -0.003 | -0.01 | -0.006 |
| | (0.009) | (0.006) | (0.01) | | (0.006) | (0.004) | (0.007) |
| 39 | -0.012 | -0.01 | -0.011 | 60 | -0.008 | -0.006 | -0.017 |
| | (0.006) | (0.006) | (0.007) | | (0.006) | (0.005) | (0.007) |
| 40 | -0.032 | -0.021 | -0.014 | | | | |
| | (0.011) | (0.007) | (0.01) | | | | |

*Note:* The table provides for each age (in years as at March 2013), differentiated by qualification group, the mean change between April 2014 and April 2013 in the probability staying on low pay when working on low pay in the three previous months ($j = 8$). For example, the probability staying on low pay declines by, on average, 10.1% points between April 2014 and April 2013 for someone who was low-paid employed in the three previous months and is 21 years old with a high qualification (Level 5-6 or higher as at March 2013).

**Table 3:** Slope of changes in low-pay persistence

| | (1) | (2) | (3) |
|---|---|---|---|
| age | 0.00038*** | 0.00055*** | 0.00886*** |
| | (0.00013) | (0.00002) | (0.00058) |
| qualification *(reference: no qualification)* | | | |
| Level 1-4 | -0.04840*** | -0.00204 | -0.20993*** |
| | (0.00819) | (0.00137) | (0.03240) |
| Level 5-6 or higher | -0.09169*** | -0.00527*** | -0.38763*** |
| | (0.00823) | (0.00137) | (0..03240) |
| qualification $\times$ age | | | |
| Level 1-4 $\times$ age | 0.00089*** | -0.00002 | 0.00300*** |
| | (0.00019) | (0.00003) | (0.00083) |
| Level 5-6 or higher $\times$ age | 0.00165*** | -0.00004 | 0.00614*** |
| | (0.00019) | (0.00003) | (0.00083) |
| N | 122 | 122 | 108 |

Columns (1) and (2) provide the respective coefficients of an OLS model using the age- and qualification differentiated average partial effects between April 2014 and April 2013. Model (3) refers to low-pay ratio in March 2018, age range is 20-55.

**Table 4:** Slope of mean wage of employer

| | March 2013 | March 2016 |
|---|---|---|
| age | -0.00016 | -0.00304*** |
| | (0.00057) | (0.00061) |
| qualification *(reference: no qualification)* | | |
| Level 1-4 | -0.00443 | 0.10949*** |
| | (0.03368) | (0.03646) |
| Level 5-6 or higher | -0.02468 | 0.25353*** |
| | (0.03368) | (0.03646) |
| qualification $\times$ age | | |
| Level 1-4 $\times$ age | 0.00039 | -.00177** |
| | (0.00080) | (0.00087) |
| Level 5-6 or higher $\times$ age | 0.00231*** | -0.00277*** |
| | (0.00080) | (0.00087) |
| constant | 7.96337*** | 8.3778*** |
| | (0.02381) | (0.02578) |
| N | 122 | 122 |

The table provides the coefficients of a simple OLS model of the mean wage of employer, differentiated according to the age and qualification-level of the low-paid worker.