Persistence of Low Pay Employment

Gail Pacheco
New Zealand Work Research Institute
(Auckland University of Technology)

Alexander Plum†
New Zealand Work Research Institute
(Auckland University of Technology)

Abstract

Several studies have shown that working in the low wage sector significantly increases the risk of staying low paid employed in the future. However, past literature has usually only considered changes in labour market status at the annual level and has not accounted for within-year changes of an individual's labour market position. Using a population-wide administrative dataset with monthly earnings information, this study shows that state dependence in low pay employment is highly heterogeneous: workers with a strong attachment to the low pay sector have a significantly higher probability staying low-paid employed compared to their low paid colleagues with a weak low pay attachment. Moreover, evidence is presented that the conventional identification strategy using annual data both under- and overestimates the persistence in low pay substantially.

Keywords: Low pay; Low pay persistence; State dependence; Administrative data

JEL classification: J62, J31

† Correspondence to: New Zealand Work Research Institute, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand. *E-mail address*: alexander.plum@aut.ac.nz

1. Introduction

Public (and academic) debate regarding rising levels of inequality has surged in recent years, often fuelled by an increasing or stagnating share of low pay employment. The average low pay incidence in the OECD in 2015 was approximately 16 percent, though with remarkable gender and country differences.[1] Related to this, the number of studies analysing the labour market prospects of low paid employed has increased noticeably, often asking to what extent low-paid employment may operate in the capacity of a 'stepping-stone' towards improved labour market outcomes versus lead into a 'no-pay – low-pay' cycle. Though the findings on this debate are rather mixed (for a discussion see Plum 2015), most studies present evidence in favour of state dependence in low pay. This means, employment in the low wage sector itself increases the risk of being employed on a low wage in the future compared to other labour market positions.

The aim of this study is to utilise monthly administrative data on wages and salary to investigate state dependence in low pay, and compare our findings with the prevailing identification strategy in relevant literature – namely, the reliance on survey data, and estimates based on earnings information for just one month within each year. This study shows that not accounting for monthly variation in wages across the year has a severe impact on the estimated size of low pay persistence. Furthermore, controlling for differences in the intensity of low pay attachment also has a considerable impact on the estimated probabilities of staying low paid employed. These findings point to a heterogeneous effect of past periods of low pay employment.

From a theoretical perspective, one major argument for state dependence in low pay employment is that low paid jobs might send a negative signal: the employer does not know the true productivity of a worker and thus might look at the labour market history of the employee as a proxy. Thus, working on low pay might be considered as working in a 'bad job' (e.g. Acemoglu 2001). Layard et al. [1991: 249] popularized this concern by the famous remark: "While unemployment is a bad signal, being in a low-quality job may well be a worse one."

To analyse the above labour market dynamics empirically, information on individual earnings and characteristics are required. Surveys (like the British Household Panel Survey (BHPS), the German Socio-Economic Panel (SOEP), and/or the Household Income and Labour Dynamics in Australia Survey (HILDA)) are often employed for these empirical exercises, as they provide a rich set of individual and labour market related information. However, one limitation of survey data is that earnings information are collected with respect to one time point in a year. Though not discussed in the low pay literature, it is implicitly assumed that earnings variation within a year does not have a major effect on estimation results. This assumption is only justified if individuals' wage variation during a year is minimal.

However, if wages are not constant within a year, the intensity with which an individual is employed in the low wage sector might also not be constant. For example, an individual might be identified as low-paid employed at the interviewed month in a survey as their wage was below the low-pay threshold. But if their wage dynamics meant that was one of only a few

[1] Data retrieved on 19 April 2018 from the OECD: http://stats.oecd.org/

months on low pay, and that the majority of the year they were earning above the threshold, then accounting for monthly variation in wages illustrates the weak attachment this individual has to the low pay sector. Likewise it is possible for someone to be classified as higher-paid employed in the interviewed month, even if that individual has spent most of the remaining months of the year in the low wage sector, and thus has a strong low-pay attachment.

This study assesses the plausibility of assuming relatively constant wages, as well as the impact when this assumption is not realised on estimates of low pay persistence. For this purpose we employ population-wide monthly administrative data on individual wages and salaries. This unit record information is sourced for the time period of 2007 to 2013 from the Integrated Data Infrastructure (IDI) in New Zealand (NZ). The IDI contains micro data from a range of government agencies and enables the researcher to join these on the individual level. For the purposes of this study, tax data from Inland Revenue are used, which covers the entirety of the NZ working population. The main advantage of these data is that accurate information are provided on a monthly basis and that survey related issues like panel attrition (e.g. Cappellari & Jenkins 2008) and measurement error associated with self-reporting do not need to be addressed. Furthermore, the high number of observations in combination with a long time period aids in providing a detailed picture of labour market movements.

In a similar vein to past literature, we focus on prime aged male workers. We initially determine the intensity of low pay attachment for each worker by calculating the number of low pay months experienced within each year relative to the total number of employed months. Based on these constructed ratios, three groups are formed: those without any low pay experience (no low pay attachment); those spending less than half of their employed months in the low pay sector (weak low pay attachment); and the final group spending at least fifty percent of their employed time in low pay (strong low pay attachment). We find a noticeable variation in low-pay attachment across the working population, with some workers on low pay for just a few months in a year, whereas others spend most of their employment in the low wage sector. Moreover, the correlation of low pay sector attachment over time appears rather stable and is substantially higher after accounting for monthly variation in earnings, compared to the traditional method of comparing wages of a single month (via survey data) across consecutive years.

To derive the impact of labour market position on future labour market outcomes, we follow a common approach in the low-pay literature and apply dynamic random-effects multinomial logit estimators. The estimation results indicate that there is substantial heterogeneity in the risk of facing low pay depending on past strength of attachment to the low pay sector. More specifically, we find that an individual has a significantly higher risk of strong low pay attachment, if they experienced strong low pay attachment in the past, compared to their counterparts who experienced weak or no low pay attachment. Furthermore, when comparing our findings to the prevailing identification strategy in the literature we find that low pay persistence is overestimated for workers with a weak low-pay attachment and underestimated for individuals with a strong low-pay attachment.

The remainder of this paper is structured as follows: Section 2 provides an overview of the current literature on low pay dynamics; Section 3 describes the NZ labour market context; Section 4 presents a conceptual framework for highlighting the effect of not accounting for monthly variation in wages; Section 5 encompasses an overview of the administrative data and key descriptives; Section 6 presents the econometric model; in Section 7 results are provided, and the final section concludes.

2. Literature Review

The number of studies on low pay has increased substantially in recent years. The aim of this section is to provide a brief overview of country specific empirical findings on low pay dynamics and prior estimates of the magnitude of low pay persistence (Table 1).

For the case of the United Kingdom, Stewart & Swaffield (1999) use the first five waves (1991-95) of the British Household Panel Survey (BHPS). Three different indicators are used to identify low-paid employed (half the median, half the mean, and two-thirds the median of gross hourly wages). Using a bivariate probit model, the authors find 'considerable persistence in low pay' [p. 40]. Stewart (2007), also uses BHPS data (for the period 1991-96), and focusses on the interrelationship between low pay[2] and unemployment. After application of various random and fixed effects specifications, he finds that 'low-wage jobs act as the main conduit for repeat unemployment' [p. 511]. In another study stemming from the use of BHPS data, Cappellari & Jenkins (2008) use an endogenous switching model, to uncover substantial persistence in low pay employment. Most recently, using data from both the BHPS and Understanding Society, Cai et al. (2017) apply dynamic multinomial logit models to estimate the extent of state dependence in low pay. In a similar fashion to the studies above, the authors conclude that 'those employees who are on low pay are more likely to be found on low pay in the future, compared with those who are (…) unemployed or on higher pay' [p. 27].

With respect to the Italian labour market, Cappellari (2007) uses data from the Survey on Households Income and Wealth (SHIW) for the waves 1993 to 2002 (noting that this is a bi-annual survey). Low paid employed are defined as the poorest fifth of the wage distribution. Multivariate probit models that control for endogeneity of initial conditions, earnings attrition and educational attainment are estimated. The author concludes that there is 'considerable state dependence: the experience of low pay raises the probability of subsequent low pay episodes' [p. 465].

Low pay persistence has also been examined in the German context. Uhlendorff (2006) uses data from the SOEP for the years 1998 to 2003 and applies two different measures to identify low-paid employed: gross wages below two thirds of the median hourly wages and the first quintile of the wage distribution. Applying a dynamic multinomial logit model, the author presents evidence for 'strong true state dependence in low pay' [p. 18] and 'a strong link between low pay and no pay' [p 18]. Mosthaf (2014) uses for the years 2000-06 data from the German Integrated Employment Biographies Sample (IEBS), which is an administrative data set with rich information on employment on a daily basis. Due to the lack of the exact working time, an individual is identified as low-paid if he earns less than two thirds of the median gross wage of all full-time employed individuals. Applying dynamic multinomial logit models with random effects the author detects noticeable state dependence in low pay.

[2] Low paid employed is defined as receiving a gross hourly wage of below £3.50, in 1997 terms.

Using data from the European Community Household Panel (ECHP), Clark & Kanellopoulos (2013) analyse the persistence of low pay employment for twelve European countries over the period 1994 – 2001. Two different measures are used for defining low-pay employment: *i*) the OECD threshold of two-thirds of the mean hourly wages and *ii*) if the individual is in the lowest three deciles of the pay distribution.[3] Using random effects probit estimators, the authors find evidence for 'positive, statistically significant state dependence in every single country' [p. 122].

Table 1: Low pay persistence of related studies

Study	$P(\mathrm{Lp}_t \vert \mathrm{Hp}_{t-1})$	$P(\mathrm{Lp}_t \vert \mathrm{Lp}_{t-1})$	$P(\mathrm{Lp}_t \vert \mathrm{ue}_{t-1})$
Uhlendorff (2006)	0.024 – 0.038	0.049 – 0.077	0.048 – 0.083
Mosthaf (2014)	0.033 – 0.007	0.091 – 0.168	0.023 – 0.076
Clark & Kanellopoulos (2013)*		UK: 0.071	
		Germany: 0.064	
		Italy: 0.045	
Cai et al. (2017) **	0.160	0.272	0.202

Note: The studies of Stewart & Swaffield (1999), Cappellari (2007) and Cappellari & Jenkins (2008) are restricting the analysis to employed. Stewart (2007) does not provide any number on low pay persistence. * 2/3 of median as provided in Table 5 [p. 127]. ** Refers to estimation results for males using the Understanding Society data.

To analyse the effect low pay employment on the future labour market position on the Australian labour market, Buddelmeyer et al. (2010) use data from Household, Income and Labour Dynamics in Australia (HILDA). In their study, a person is defined as low paid if their wage is bellow two-thirds of the median gross hourly wage. Dynamic random effects probit models are applied and the author show that being low-paid employed increases the risk of future unemployment; however, these findings are only significant among women. Fok et al. (2014) also use data from HILDA to estimate transitions between unemployment, low-pay and higher pay employment. To identify low paid employed, a person must have an hourly rate of pay below 120% of the hourly minimum wage and a weekly wage below 120% of the weekly minimum wage. The authors use dynamic multinomial logit model with random effects to estimate state dependence in low pay. One finding is that there is sizeable state dependence in low pay.

3. The NZ low pay sector

There is limited empirical analyses regarding the NZ low pay sector. There is one descriptive study by Cochrane et al. (2018) that use two different definitions to identify low paid employed: the first one accounting for the relative distance to the median wage (labelled as OECD low pay) and the second one in relation to the minimum wage (labelled as 120% MW low pay). As a result of these different measures the low pay incidence varies; in 2015 about 11.1 percent of the employed were low paid according to the OECD definition and 24.9 according to the minimum wage definition.

To derive the comparability of the NZ labour market with others countries from the OECD, some OECD economic and labour market related indicators for Australia, Germany, New Zealand, Great Britain and United States are presented in Table 2. It can be seen that evaluated by

[3] For robustness reasons various different thresholds are applied.

their contribution to GDP the main four sectors are trade, public administration, industry and real estate activities: their summed percentage contribution to GDP is about 64 percent. This is on a comparable level with the respective contribution of GBR (64 percent) Australia (64 percent) and the US (67 percent). Due to a substantially greater industry sector, the percentage of GDP generated by those four sectors is noticeably greater for Germany (71 percent).

Table 2: Economic and labour market related indicators for five selected OECD countries

	AUS	DEU	NZL	GBR	USA
Gross value added at basic prices (2014, ISIC rev4)					
Agriculture, forestry and fishing	2.5%	0.8%	6.8%	0.7%	1.3%
Industry, including energy	16.6%	25.9%	15.4%	14.0%	16.9%
Construction	8.6%	4.5%	6.3%	6.0%	4.0%
Distributive trade, repairs *	16.7%	15.9%	17.7%	18.0%	15.9%
Information and communication	2.9%	4.7%	3.0%	6.0%	5.9%
Financial and insurance activities	9.1%	4.1%	6.3%	7.5%	7.3%
Real estate activities	12.3%	11.0%	14.8%	13.5%	12.3%
Prof., scientific, techn.; admin., **	10.6%	10.9%	10.4%	12.1%	11.5%
Public admin.; compulsory s.s.; ***	17.9%	18.0%	15.8%	18.3%	21.8%
Other service activities	2.9%	4.1%	3.5%	4.0%	3.2%
Low Pay Incidence (men, 2015)	13.5%	15.7%	12.8%	15.5%	21.2%
Minimum wage relative to median wages of full-time workers (2016)	53.8%	46.7%	60.5%	49.0%	34.9%
Strictness of employment protection – individual and collective dismissals (regular contracts, 2013)	1.94	2.84	1.01	1.66	1.17
Collective bargaining coverage (2010)	58.6%	59.8%	15.6%	30.9%	12.6%

Source: OECD Statistics (2018). * includes transport; accommod., food serv.; ** includes support serv. Activities; *** includes education; human health

Looking at labour market indicators (bottom part of Table 2), it can be seen that New Zealand has a low pay incidence (13 percent) which is on a comparable level as Australia (14 percent), UK (16 percent) and Germany (16 percent). The US faces the highest share of low pay incidence (21 percent). One explanation for the slightly smaller size of the low pay sector could be the relatively high level of the minimum wage: set into ratio with the median wage of full-time workers, the level is 61 percent for New Zealand, 54 percent for Australia, 49 percent for UK, 47 percent for Germany and 35 percent for the US.

Moreover, the OECD data show that New Zealand has a low level of strictness of employment protection, which is on a comparable level as the US. Furthermore, the collective bargaining coverage is little and also comparable with the US level.

4. Conceptual framework

4.1 Basic concept

To fix ideas and to provide an impression of the influence of monthly variation of wages and salaries, we start with a model that is used for describing earning dynamics (e.g. Baker & Solon 2003, Cappellari & Jenkins 2014):

$$Y_{ik_m} = \mu_k + y_{ik_m} \tag{1}$$

with Y_{ik_m} referring to wages of individual $i = 1, \dots, N$ in year $k = 1, \dots, K$ at month $m = 1, \dots, 12$.[4] These wages consist of two parts: μ_k, which is unrelated to the individual, and an individual-specific deviation y_{ik_m}. The second term itself can be again separated into two parts:

$$y_{ik_m} = \alpha_i + \nu_{ik_m} \tag{2}$$

The first term refers to a random individual-specific time-invariant component and the second term is a month specific idiosyncratic shock. For simplification, it is assumed that α_i and ν_{ik_m} are orthogonal to each other and ν_{ik_m} is serially uncorrelated. The variance takes the following form:

$$Var(y_{ik_m}) = \sigma_\alpha^2 + \sigma_\nu^2 \tag{3}$$

An individual is identified as being low paid in month m if the monthly wages are below the threshold τ:

$$LP_{ik_m} = \mathbf{1}(Y_{ik_m} \leq \tau) \tag{4}$$

Turning to the annual level a marker that indicates whether an individual has experienced at least one month of low pay can be written the following (with $M_{ik} \in \{1, \dots, 12\}$ referring to the individual number of employed months):

$$LP_{ik} = \mathbf{1}\left(\sum_{m=1}^{M_{ik}} LP_{ik_m} > 0\right) \tag{5}$$

Thus the annual share of individuals with a positive number of low pay employment months is $LP_k = \frac{\sum_{i=1}^{N} LP_{ik}}{\sum_{i=1}^{N} M_{ik}}$. If there is no monthly variation in the wages and salaries than this ratio is equal to the ratio of any month $\overline{m} \in \{1, \dots, M_{ik}\}$:

$$LP_k = \frac{\sum_{i=1}^{N} LP_{ik_{\overline{m}}}}{N_k} \text{ if } \sigma_\nu^2 = 0$$

Furthermore, the individual share of low paid employed months' can be derived:

$$LP_{ik}^s = \frac{\sum_{m=1}^{M_{ik}} LP_{ik_m}}{M_{ik}} \text{ with } LP_{ik}^s \in \left\{0, {}^{1}/_{M_{ik}}, \dots, 1\right\} \tag{6}$$

[4] Note that the original models account for variation of the wages on the annual and not on the monthly level.

This ratio might be understood as a marker on low pay attachment and a more aggregated version is used later. If there is no monthly variation than this marker is equal to the low pay intensity of any month $\left(LP_{ik}^{s}=LP_{ik_{\overline{m}}} \text{ if } \sigma_v^2=0\right)$.

4.2 Correlation over time

To derive how this affects the correlation over time between year $k-1$ and k, some assumptions are made for simplification. First, it is assumed that all individuals work the same amount of months in the two consecutive periods, thus $M_{ik-1}=M_{ik}=M\ \forall\ i$. Furthermore, it is assumed that the number of observations N is sufficiently large for asymptotics. To identify the low-paid employed, a relative threshold, e.g. the lowest tenth percentile, is used that is constant over time, thus $\tau_{k-1_1}=\cdots=\tau_{k-1_M}=\cdots=\tau_{k_M}=\tau$. Therefore, $\sum_i LP_{ik}^{s}=\sum_i LP_{ik_{\overline{m}}}=N\times\tau$ with month $\overline{m}\in\{1,\dots,M\}$ and this finding is independent of σ_v^2.

Note that if $\sigma_v^2=0$ than $\sum_i(LP_{ik}^{s})^2=\sum_i(LP_{ik-1}^{s}LP_{ik}^{s})=N\times\tau$ as $LP_{ik}^{s}\in\{0,1\}$; if $\sigma_v^2\to\infty$ than $\sum_i(LP_{ik}^{s})^2=\sum_i(LP_{ik-1}^{s}LP_{ik}^{s})=N\times\tau^2$ as $LP_{ik}^{s}\xrightarrow[\sigma_v^2\to\infty]{}\tau$. However, $\sum_i LP_{ik_{\overline{m}}}=N\times\tau$ independent of σ_v^2 but $\sum_i\left(LP_{ik-1_{\overline{m}}}LP_{ik_{\overline{m}}}\right)=\sum_i(LP_{ik}^{s})^2$ if $\sigma_v^2=0$ and $\sum_i\left(LP_{ik-1_{\overline{m}}}LP_{ik_{\overline{m}}}\right)=N\times\tau^2$ as $\sigma_v^2\to\infty$. The correlation coefficient takes the following form for the marker that accounts for monthly variation of wages

$$corr[LP_{ik}^{s},LP_{ik+1}^{s}]=\frac{N(\sum_i LP_{ik-1}^{s}LP_{ik}^{s})-(\sum_i LP_{ik-1}^{s})(\sum_i LP_{ik}^{s})}{\sqrt{[N\sum_i(LP_{ik-1}^{s})^2-(\sum_i LP_{ik-1}^{s})^2][N\sum_i(LP_{ik}^{s})^2-(\sum_i LP_{ik}^{s})^2]}}$$

and for the marker that compares the labour market position of month $\overline{m}$ of two consecutive years

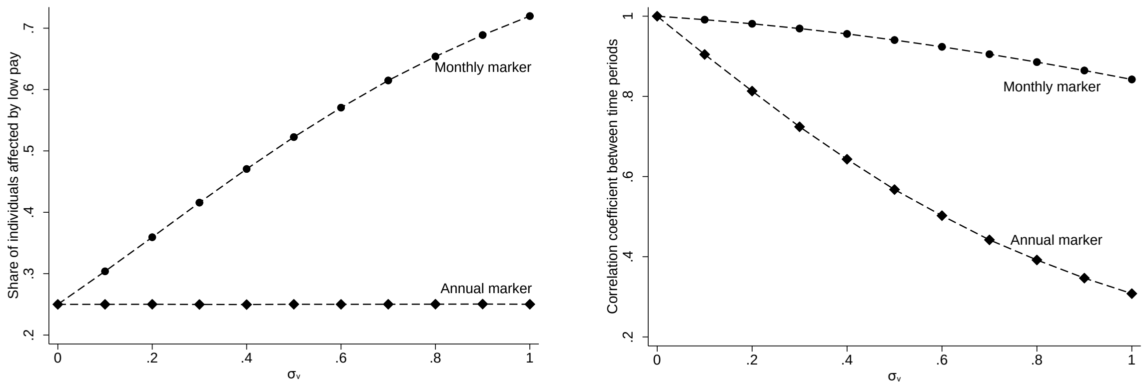
$$corr\left[LP_{ik-1_{\overline{m}}},LP_{ik_{\overline{m}}}\right]=\frac{N\left(\sum_i LP_{ik-1_{\overline{m}}}LP_{ik_{\overline{m}}}\right)-\left(\sum_i LP_{ik-1_{\overline{m}}}\right)\left(\sum_i LP_{ik_{\overline{m}}}\right)}{\sqrt{\left[N\sum_i\left(LP_{ik-1_{\overline{m}}}\right)^2-\left(\sum_i LP_{ik-1_{\overline{m}}}\right)^2\right]\left[N\sum_i\left(LP_{ik_{\overline{m}}}\right)^2-\left(\sum_i LP_{ik_{\overline{m}}}\right)^2\right]}}$$

The correlation coefficient of both markers are restricted between 1 ($\sigma_v^2=0$) and 0 ($\sigma_v^2\to\infty$) and thus correlation over times declines if monthly variation of wages increases ($\sigma_v^2>0$). However, due to $N\to\infty$ we know that $\left|\frac{\partial(\sum_i LP_{ik-1}^{s}LP_{ik}^{s})}{\partial\sigma_v^2}\right|<\left|\frac{\partial\left(\sum_i LP_{ik-1_{\overline{m}}}LP_{ik_{\overline{m}}}\right)}{\partial\sigma_v^2}\right|$. Moreover, as $\sum_i\left(LP_{ik-1_{\overline{m}}}\right)^2\perp\sigma_v^2$ and $\left|\frac{\partial\sum_i(LP_{ik}^{s})^2}{\partial\sigma_v^2}\right|>0$ we can conclude that $\left|\frac{\partial\left(corr[LP_{ik}^{s},LP_{ik+1}^{s}]\right)}{\partial\sigma_v^2}\right|<\left|\frac{\partial\left(corr\left[LP_{ik-1_{\overline{m}}},LP_{ik_{\overline{m}}}\right]\right)}{\partial\sigma_v^2}\right|$.

4.3 Simulation

To emphasize the findings a simulation is run based on 5,000 individuals of whom all were employed for 12 months. It is assumed that $Y_{ik_m}=2{,}000+200\alpha_i+200v_{ik_m}$ with $k\in\{1,2\}$, $\alpha_i=(0,1)$ and $\alpha_i=(0,x)$ with $x\in\{0,.1,\dots,1\}$. An individual is identified as being low pay if the log wages belong to the lowest 25th percentile within the respective month. 500 replications are chosen for each level of σ_v. Afterwards, the share of individuals who have experienced at least one month of low pay employment and the correlation of the low pay intensity, which is the share of low pay employment months divided by the total number of months, are calculated.

Figure 1: Simulation results (averages over 500 draws)



Source: own simulation (see numbers in text). The left panel shows the share of individuals affected by low pay when accounting for monthly variation (‘Monthly marker’) and when only considering the labour market position at one month of a year (‘Annual marker’). The right panel shows the correlation between two time points of the two marker.

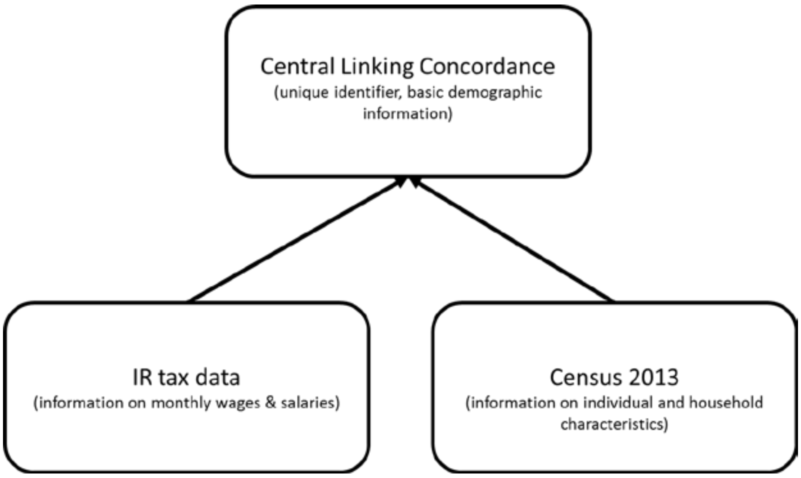
To highlight the difference between the prevailing identification strategy when using survey data which only provide information for one month in a year and the case when considering all time points, two different marker are generated. The first marker uses all monthly information provided to calculate the share of individuals affected by low pay and is labelled ‘Monthly marker’. The second marker only uses the information of the first month of the year and is called ‘Annual marker’. As depicted in Figure 1, both marker only derive identical results when there is no monthly variation of the wages. However, with an increase of the variation of the monthly wages the difference between the calculated shares of individuals who experienced low pay in the year is increasing (left panel). Furthermore, the correlation of the low-pay position between two time points is underestimated when using information that refer only to one month of a year (right panel).

5. Data and Descriptive Statistics

To analyse state dependence in low pay employment, data from Statistics New Zealand’s Integrated Data Infrastructure (IDI) are used.[5] The IDI links longitudinal microdata about individuals, households, businesses and organisations from various sources; e.g. Ministry of Business, Innovation and Employment on migration, Ministry of Education on secondary and tertiary education or the Ministry of Social Development on benefit. The backbone of the IDI is the Central Linking Concordance (CLC), which contains a list of all individuals including an assigned unique identifier for an individual along with their basic demographic information on sex, date of birth and ethnicity indicators. The identifier enables the researcher to link all the different datasets.

[5] Please note the disclaimer in the Appendix

Figure 2: The data sources



Source: own representation.

The IDI also consolidates information on person tax data from Inland Revenue. The IR tax data are provided from 1 April 1999 onwards and the geographic coverage refers to all New Zealand. Data are collected and supplied monthly to the IDI. To identify gross wages paid to the employee, two variables are used:

- First one provides information on total wages before tax for each employer.
- Second gives a code representing the source of income: i) wages & salary, ii) withholding payment, iii) benefits, iv) student allowance, v) paid parental leave, vi) pensions (superannuation) and vii) claimants compensations.

Another dataset we are using is the Census 2013. The Census was conducted on Tuesday, 5 March 2013 and contains a range of individual and household related information on all individuals living on the respective day in New Zealand. For the analysis, all three data sets are combined (see Figure 2).

For our analysis we use the gross wages before tax that come from wages and salaries. As an employee might be holding multiple jobs or changes jobs within a month, there could be more than one line item entry per month per individual. As we are considering monthly wage changes, we aggregate all monthly wages. To control for individual characteristics we use information provided by the Census 2013 and we only consider the years 2007 to 2013 for our analysis to ensure that no substantial changes on the individual level occurred. Furthermore, we restrict our sample to male employee of age 25 to 45 in 2007. We do this for several reasons:

- We do not have any information on the actual individual working hours. However, according to data from the OECD, in the respective time frame about 95 percent of those employed prime aged men are working fulltime. Moreover, also those months are excluded in which the individual was earning per month below 30 hours times the respective minimum wage times 4.2 weeks.
- The age restrictions helps to mitigate the influence of schooling or early retirement schemes.

To ensure a sufficient labour market attachment, we restrict our sample to those individuals who were employed at least for five months per year and have tax entries for the whole covered time period. This provides us with a balanced sample. To identify those worker who are low paid employed age related[6] low pay thresholds are calculated. We use as a threshold two thirds of the median monthly wages of the respective age group (OECD 1997).[7]

After identifying the intensity an individual is attached by low wages, the following three groups are generated (this identifier is called *Monthly marker*):

- *No low-pay months*: individuals without any low pay months within a year or individuals who were employed each month of a year and were observed working in the low wage sector for only one month
- *Low-pay month:* <50 *percent*: individuals who have worked in the low wage sector but less than half of their annual employment duration.
- *Low-pay month:* ≥ 50 *percent*: individuals who have worked at least half of their total annual employment period in the low wage sector.

Afterwards, we draw a random subsample that consists of 77 250 observations. To get an impression of the labour market attachment, the relative group size of the respective groups is listed in Table 3 (last column). According to this definition, about every fifth worker has some kind of attachment to the low pay sector. Furthermore, the two groups that should reflect the intensity of being attached to the low pay sector are of comparable size, with a slightly bigger share with a weak low-pay attachment.

Table 3: Comparing the identification of the low pay employed [a]

		Annual Marker		
		Higher pay$_t$	*Low pay*$_t$	Share$_t$
Monthly marker	*No low-pay months*$_t$	97.46	2.54	82.4
	Low-pay month:$<50\%_t$	70.92	29.08	9.72
	Low-pay month:$\geq 50\%_t$	26.86	73.14	7.88
	Share$_t$	89.32	10.68	

Source: IDI data (2018), own calculations. *N*=77 250.

To demonstrate the difference to the prevailing identification strategy a second marker is constructed – called *Annual Marker* – which refers to the labour market position in the first observed month of a year.[8] Here, individuals are differentiated between higher-paid employed (Hp) and low-paid employed (Lp). The cross tabulation in Table 3 indicates, that about 70 percent of those individuals who were identified as having a weak level of low pay attachment are identified as higher paid employed – and three quarter of those who were considered as having a high level of low pay attachment are grouped into the low pay category. Moreover, when applying the Annual Marker the size of individuals with low-pay attachment half's to about ten percent.

To derive a more complete picture of the intertemporal correlation, the correlation structure for the period 2008 to 2013 is presented in Figure 3. The figure is very comparable to what was

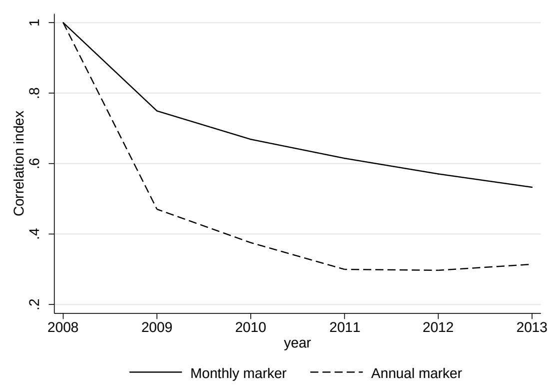
[6] Following age groups are considered: <30, 30-35, 35-40 and 40+.

[7] Findings hold for the case when using as indicator if the earnings of the employed belongs to the bottom 25th percentile.

[8] Very similar results are derived when using a randomly assigned month.

derived in the simulation, indicating that not accounting for the monthly variation of wages might underestimate the correlation of the attachment of low pay. For example, the Annual Marker gives a correlation coefficient of about 0.31 for the labour market position between 2008 and 2013 – in the case of the Monthly Marker this is nearly two times greater and ranges above 0.53.

Figure 3: Intertemporal correlation of the Monthly and Annual Marker



Source: IDI data, own calculations. *N*=77 250.

Another possibility to derive descriptively the correlation is by generating transition matrix of the labour market position. This matrix provides the probability of being in one labour market position at time point t conditional on the labour market position at $t-1$. As can be seen from Table A1, those individuals without any low pay experience within a year have a very high conditional probability of about 95 percent to experience no low pay month in the preceding year. Referring to those individuals with a weak low pay attachment the conditional probabilities of having no or only a limited low pay experience in the next year take comparable sizes of above 40 percent. Furthermore, those individuals with a strong attachment to the low pay sector have a conditional probability of about 74 percent to experience a high number of low pay months. Finally, transitions from one to the other extreme are found to be rare: for example, the share of individuals moving from no low pay attachment to a strong on is below one percent, the opposite direction about six percent. Looking at the transition matrix when applying the Annual Marker, one obvious difference is the share of individuals who move from low pay to higher pay, which is about 48 percent, and another aspect is the low share of individuals who stay low paid, about 51 percent (Table A2).

To sum up the findings, the descriptive statistics indicate that the intensity a worker is attached to the low pay sector is rather heterogeneous among the workforce. Moreover, the intensity an individual is attached to the low pay sector is found to be rather stable over time.

6. Econometric Model

In general, when a dynamic model is applied, it must address several aspects, such as unobserved heterogeneity (Heckman 1981a) and its correlation with the initial conditions (Heckman 1981b). As Skrondal and Rabe-Hesketh (2014) have pointed out, not accounting for these aspects might cause biased results.

In this study, we apply a dynamic random effects multinomial logit model which was also used in various other low pay studies (e.g. Uhlendorff 2006, Mosthaf 2014, Fok et al. 2015, Cai et al. 2017). In this study we consider three labour market positions and j is equal to 1 if the individual has no low pay months, 2 if having low pay months of below 50 percent and 3 if having low pay months of at least 50 percent. Thus, the probability of individual $i \in \{1, \ldots, N\}$ to be in the labour market state y_{it} at time point $t \in \{1, \ldots, T\}$ can be written as:

$$P(y_{it} = j | y_{it-1}, X_i, \alpha_{ij}) = \frac{exp(X_i' \beta_j + y_{it-1}' \gamma_j + \alpha_{ij})}{\sum_{k=1}^{3} exp(X_i' \beta_k + y_{it-1}' \gamma_k + \alpha_{ij})} \tag{7}$$

Note that X_i refers to a vector of explanatory variables[9] and y_{it-1} is a vector of dummy variables with respect to the lagged labour market position. Moreover, a time-invariant error terms α_{ji} is included to capture individual-specific effects like motivation or ability. To identify the model we chose that the category with having no low pay month as reference category and therefore the coefficient vectors β_1, γ_1 and α_{i1} are set equal to zero.

However, the labour market position in the initial period might not be randomly distributed, due to a correlation between the time-invariant error term and the initial conditions.[10] To take care of the "initial conditions problem", we follow the suggestion of Wooldridge (2005) by applying a conditional random-intercept model:[11]

$$\alpha_{ji} = y_{i0}' \lambda + \kappa_{ji} \tag{8}$$

It is assumed that the random effects are normally distributed $\kappa_{ji} \sim N\left(0, \sigma_{\kappa_j}^2\right)$ and are correlated by ρ_κ. After substituting equation (8) into (7) the likelihood function for individual i takes the following form:

$$L_i = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \prod_{j=2}^{3} \left\{ \frac{exp(X_i' \beta_j + y_{it-1}' \gamma_j + y_{i0}' \lambda + \kappa_{ji})}{1 + \sum_{j=2}^{3} exp(X_i' \beta_j + y_{it-1}' \gamma_j + y_{i0}' \lambda + \kappa_{ji})} \right\}^{d_{ijt}} f(\kappa) d\kappa \tag{9}$$

Note that d_{ijt} equals 1 if individual i is in state j at time point t and zero otherwise. To integrate out the random effects, we use maximum simulated likelihood. Using random numbers based on prime numbers (also called Halton draws, see Train 2009), two times R standard uniform distributed draws are derived and transformed by the inverse cumulative standard normal distribution. For each draw, the likelihood is derived for each observation, multiplied over all individuals and time-points and finally averaged over all draws (using 50 draws):

$$MSL = \prod_{i=1}^{N} \frac{1}{R} \sum_{r=1}^{R} \left\{ \prod_{t=1}^{T_i} P_{it}(\kappa_1^r, \kappa_2^r) \right\} \tag{10}$$

[9] Note that as we use data from the Census 2013 and age group specific estimations are conducted no information on the individual or household level are time-varying.

[10] As a robustness estimation we also restricted the sample to those individuals who were on low-pay in 2007. However, this not affect the results.

[11] In Wooldridge (2005) it is suggested to include the time means of the explanatory variables, however as we only consider time constant explanatory variables this aspect is dropped.

7. Results

7.1 Estimation results

Aim of this study is to analyse persistence in low pay employment. The existing literature is extended by controlling for the individuals' low pay attachment intensity. To capture the degree of low pay attachment, the individuals are differentiated into three groups. With respect to the econometric model we follow the standard approach in the economic literature and apply a dynamic multinomial random effects logit model.

Table 4: Regression results on low pay persistence

	Low-pay month: < 50 *percent*$_t$	*Low-pay month:* ≥ 50 *percent*$_t$
Low-pay attachment at *t-1*		
No low-pay months$_{t-1}$	*ref. category*	
Low-pay month:< 50 *percent*$_{t-1}$	1.358	2.416
	(0.051)	(0.096)
Low-pay month:≥ 50 *percent*$_{t-1}$	2.910	5.643
	(0.088)	(0.123)
Low-pay attachment at *t=0*		
No low-pay months$_{t=0}$	*ref. category*	
Low-pay month:< 50 *percent*$_{t=0}$	1.956	2.548
	(0.073)	(0.130)
Low-pay month:≥ 50 *percent*$_{t=0}$	2.578	4.335
	(0.109)	(0.201)
exogenous variables†	✓	✓
$\sigma^2_{\kappa_1}$	2.135	
	(0.118)	
$\sigma^2_{\kappa_2}$	4.390	
	(0.366)	
ρ_κ	0.875	
	(0.019)	
Log Likelihood	-24 663.309	
N	77 250	

Note: IDI (2018) and own calculations. Numbers in parenthesis refer to standard errors. † The following explanatory variables are included: disability, only English speaking household, ethnicity (European, Maori, Pacific, Asian), post-school qualification (4 categories), year, having non-employment months in the previous year (dummy).

Estimation results are presented in Table 4, with being employed without any low pay month as the reference category. Displayed are the coefficients of the lagged labour market positions and the size and correlation of the random effects coefficients. With respect to the unobserved heterogeneity, the variances of the random effect error terms are of noticeable size and highly significant. Moreover, evidence for a positive correlation of the random effects is presented, indicating that an individual who is more likely of becoming low paid employed for less than half of the employment months is also more likely to become low paid employed for most of the time in a year instead of experiencing no low pay months.

Referring to the coefficients that are capturing the effect of the initial labour market position, we find indications that those with a weak low pay attachment in 2007 are either more likely to stay in this labour market position or become strongly attached to the low pay sector compared to someone without any low pay experience in 2007. The same pattern can be found for those individuals who were strongly attached to the low-pay sector in 2007, though on a much more pronounced level.

Furthermore, those coefficients related to the lagged labour market position provide evidence for persistence in the low-pay intensity: compared to individuals without any low-pay attachment at $t-1$, being on low-pay on a low $\left(\hat{\beta}_{Lp\ spell:<50\%_{t-1}} = 1.358\right)$ or high intensity $\left(\hat{\beta}_{Lp\ spell:\geq 50\%_{t-1}} = 2.910\right)$ increases significantly the chances of being weakly attached to the low pay sector in the subsequent period. Moreover, both coefficients are significantly different from each other at the 1 percent level. Referring to the risk of being low paid employed for the majority of the employed months, the same pattern can be found, though the coefficients are of much greater size $\left(\hat{\beta}_{Lp\ spell:<50\%_{t-1}} = 2.416, \hat{\beta}_{Lp\ spell:\geq 50\%_{t-1}} = 5.643\right)$.

7.2 Persistence in low pay

In a next step, the probability of being in one of the three labour market positions in dependence of the previous labour market position is calculated (see Table 5, first column). As shown, the mean probability of an individual to not experience a low pay month who were not on low pay at $t-1$ is about 93 percent; this number declines to 81 percent if the individual had a weak low-pay attachment in $t-1$ and to 51 percent in the case of a strong low-pay attachment. Beside of a decline of the probability an increase of the standard deviation can be found, indicating a heterogeneous effect of the past low pay attachment.

Table 5: Predicted probabilities of low pay persistence

	Total	***No lp months$_{t=0}$***	***Lp month: $<50\%_{t=0}$***	***Lp month: $\geq 50\%_{t=0}$***
$P(No\ lp\ months_t \mid No\ lp\ months_{t-1})$	0.933	0.978	0.838	0.709
	(0.098)	(0.010)	(0.065)	(0.093)
$P(Lp\ months: < 50\%_t \mid No\ lp\ months_{t-1})$	0.060	0.021	0.152	0.238
	(0.082)	(0.010)	(0.061)	(0.075)
$P(Lp\ months: \geq 50\%_t \mid No\ lp\ months_{t-1})$	0.008	0.001	0.010	0.054
	(0.018)	(0.000)	(0.005)	(0.021)
$P(No\ lp\ months_t \mid Lp\ months: < 50\%_{t-1})$	0.812	0.917	0.558	0.335
	(0.209)	(0.035)	(0.110)	(0.099)
$P(Lp\ months: < 50\%_t \mid Lp\ months: < 50\%_{t-1})$	0.145	0.075	0.370	0.406
	(0.139)	(0.031)	(0.091)	(0.059)
$P(Lp\ months: \geq 50\%_t \mid Lp\ months: < 50\%_{t-1})$	0.042	0.008	0.073	0.259
	(0.080)	(0.004)	(0.025)	(0.061)
$P(No\ lp\ months_t \mid Lp\ months: \geq 50\%_{t-1})$	0.514	0.636	0.147	0.041
	(0.245)	(0.102)	(0.059)	(0.020)
$P(Lp\ months: < 50\%_t \mid Lp\ months: \geq 50\%_{t-1})$	0.253	0.231	0.421	0.222
	(0.084)	(0.059)	(0.046)	(0.038)
$P(Lp\ months: \geq 50\%_t \mid Lp\ months: \geq 50\%_{t-1})$	0.233	0.133	0.432	0.736
	(0.205)	(0.048)	(0.066)	(0.050)

Note: IDI (2018) and own calculations. Numbers in parentheses refers to the standard deviation.

Referring to the risk of becoming low-paid employed, those without any low-pay attachment have a very little risk (on average <1 percent) of becoming strongly attached to the low-pay sector in the subsequent period. A small risk can also be found for those who were weakly attached to the low-pay sector (on average <5 percent for strong low pay attachment). Unsurprisingly, the highest risk is to be found when being strongly attached to the low-pay sector:

the mean probability being also strongly attached to the low pay sector in the subsequent period is about 23 percent.

Table 6: Distribution of the initial labour market condition

	No low-pay months$_t$	*Low-pay month: $< 50\%_t$*	*Low-pay month: $\geq 50\%_t$*	Share$_{t=0}$
No low-pay months$_{t=0}$	87.5	8.86	3.64	82.4
Low-pay month:$< 50\%_{t=0}$	43.48	32.55	23.96	9.72
Low-pay month:$\geq 50\%_{t=0}$	13.18	20.49	66.33	7.88
Share$_t$	77.37	12.08	10.56	

Source: IDI data (2018), own calculations. *N*=77 250.

As we have seen in Table 4, the coefficients referring to the initial labour market position have a strong impact. Moreover, in the data a strong correlation of the initial labour market position with its subsequent ones can be observed (see Table 6). For example, two third of those individuals who were initially low-paid employed for at least 50 percent of the employed months are in this labour market position later on. Thus, the degree of low pay persistence is differentiated according to the initial labour market positions (Table 2, columns 2-4). The following findings can be listed:

1. A rather small effect can be found for those without any low pay experience in the previous period. Though experiencing a strong low pay attachment in the initial period leads to a decline in the probability of experiencing no low pay month, the risk of becoming strongly attached to the low pay sector only increases gradually (from <1 percent to about 5 percent).
2. Moreover, those individuals who were weakly attached to the low pay sector in the previous period also experience a decline in the probability of exiting the low pay sector in dependence of the degree of low-pay attachment in the initial period. But the increase of the risk being strongly attached to the low pay sector is still moderate (from <1 percent to about 26 percent) and this risk is exceeded by the chances of exiting the low pay sector (34 percent).
3. Referring to those with a strong low pay attachment in the previous period, differences are noticeable. On the one side, the chances exiting the low pay sector declines substantially (from 64 percent to 4 percent). Likewise, the risk being strongly attached to the low pay sector increases, e.g. those who were initially strongly attached to the low pay sector and were also strongly attached in the previous period have in average a probability of about 74 percent keeping that attachment level.

7.3 Low pay persistence according to prevailing identification strategy

In the following, the identification strategy of the annual marker which should reflect the prevailing identification strategy is used to re-estimate the low pay labour dynamics. In contrast to the previous regression model, the individuals are only differentiated into low-paid and higher-paid employed according to their labour market position in the first month they have been employed for each year. As there are no only two different labour market position that are considered, a random effects logit model is applied (see Table A3). In line with the previous findings, the regression results indicate that being on low pay in the previous period increases significantly the chances on being on low pay in the subsequent period compared to when being higher paid. Moreover, being low-paid employed in the initial period increases the risk of experiencing low pay employment in the future.

In the following, probabilities in dependence of the previous labour market position are calculated (see Table A4). In general, the probability being affected by low-pay is rather small, though increases from about 5 to 12 percent when being low-paid instead of higher-paid employed. Differentiating the probabilities according to the initial labour market position provides a more differentiated picture of low-pay persistence. However, those who were on low-pay in the initial and previous have on average a chance of 58 percent of moving into a higher paid job. Thus, using the annual marker would indicate much better prospects of exiting the low wage sector than derived on using monthly employment information.

8. Conclusion

Numerous prior studies have estimated the extent to which individuals who were working in the low wage sector at time point $t-1$ are likely to find themselves again in this labour market position in the subsequent period t. Based usually on survey data, and a reliance on wage levels at one time point per year, past evidence has generally concluded that compared to their higher-paid colleagues, low-paid employed face a significantly higher risk of being low-paid employed in the future.

Given that wages may not be constant over a year, from an individual perspective, the respective labour market position at one point of time may not reflect the actual intensity with which a person is attached to the low pay sector. To overcome this shortcoming associated with survey information, this study utilises population-wide administrative data with monthly information on wages. These data permit differentiation of those employed into three groups: those without any low pay experience (no low pay attachment); those spending less than half of their employed months in the low pay sector (weak low pay attachment); and the final group spending at least fifty percent of their employed time in low pay (strong low pay attachment).

To estimate the size of state dependence in low pay, we use a dynamic random-effects multinomial logit model. We find a noticeable persistence in low pay, especially when differentiating according to the initial intensity of low pay attachment: for example an initially strong attached worker who was also strongly attached at $t-1$ has on average a probability of about 74 percent of maintaining that strong attachment. Additionally, the probability of an individual in the strong attached group exiting the low-pay sector without any low-pay attachment is estimated to be 4 percent. Conversely, we also find that individuals with no low-pay experience are very unlikely to form a strong attachment to the low wage sector in future time periods.

When using the prevailing identification strategy, the heterogeneity of past low-pay cannot be detected at that granularity. This has two effects on the estimated size of low pay persistence: on the one hand, low pay persistence is overestimated for those with a weak past low pay attachment and underestimated for those with a strong past low pay attachment. These findings underline the necessity to control for the intensity of low pay attachment.

Furthermore, based on the prevailing identification method, we would conclude a high level of state dependence in low pay, and interestingly, an even higher probability of low paid workers climbing up the wage ladder to higher pay in future time periods. Thus, if policy design were based on these findings one might conclude that low-paid employment offers a 'stepping-stone' to higher-paid employment. However, we find that after accounting for the level of attachment to the low wage sector (based on monthly information on wages) those with a strong attachment have very little chance of exiting this sector. This put strong doubts on whether there is a 'stepping-stone' effect of low pay and whether 'any job is helpful' with respect of climbing up the wage ladder. This finding aligns with the hypothesis that not every job contributes to the individuals' human capital level.

References

Acemoglu, D. (2001). Good jobs versus bad jobs. *Journal of Labor Economics*, *19*(1), 1-21.

Buddelmeyer, H., Lee, W. S., & Wooden, M. (2010). Low-Paid Employment and Unemployment Dynamics in Australia. *Economic Record*, 86(272), 28-48.

Cai, L., Mavromaras, K. & Sloane, P. (2017), Low Paid Employment in Britain: Estimating State-Dependence and Stepping Stone Effects. *Oxford Bulletin of Economics and Statistics* doi:10.1111/obes.12197

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

Cappellari, L., & Jenkins, S. P. (2008). Estimating low pay transition probabilities accounting for endogenous selection mechanisms. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(2), 165-186.

Clark, K., & Kanellopoulos, N. C. (2013). Low pay persistence in Europe. *Labour Economics*, 23, 122-134.

Cochrane, B., Fletcher, M., Pacheco, G. & Plum, A. (2018). *Low pay in NZ.* A report commissioned by the Ministry of Business, Innovation & Employment to the New Zealand Work Research Institute, Auckland University of Technology (NZWRI, AUT).

Fok, Y. K., Scutella, R., & Wilkins, R. (2015). The Low‐Pay No‐Pay Cycle: Are There Systematic Differences across Demographic Groups? *Oxford Bulletin of Economics and Statistics*, *77*(6), 872-896.

Heckman, J. J. (1976). A life-cycle model of earnings, learning, and consumption. *Journal of political economy*, 84(4, Part 2), S9-S44.

Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in Labor Markets* (pp. 91-140). University of Chicago Press.

Heckman, J.: 1981b, The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process, in C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.

Layard, P., Nickell, S., & Jackman, R. (1991). *Unemployment: macroeconomic performance and the labour market*. Oxford: Oxford University Press.

Mosthaf, A. (2014). Do Scarring Effects of Low‐Wage Employment and Non‐Employment Differ BETWEEN Levels of Qualification? *Scottish Journal of Political Economy*, 61(2), 154-177.

Skrondal, A., & Rabe‐Hesketh, S. (2014). 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)*, 63(2), 211-237.

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

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

Uhlendorff, A. (2006). *From no pay to low pay and back again? A multistate model of low pay dynamics*. IZA Discussion Papers, No. 2482, http://nbn-resolving.de/urn:nbn:de:101:1-2008051370

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

Appendix

Table A1: Transition matrix of the labour market positions (Monthly marker)

	No lp months$_t$	*Lp month:$< 50\%_t$*	*Lp month:$\geq 50\%_t$*	*Total$_{t-1}$*
No lp months$_{t-1}$	94.98	4.59	0.43	81.45
Lp month:$< 50\%_{t-1}$	44.69	41.77	13.54	10.20
Lp month:$\geq 50\%_{t-1}$	5.72	20.67	73.61	8.35
Total$_t$	82.4	9.72	7.88	

Source: IDI data, own calculations. *N*=77 250.

Table A2: Transition matrix of the labour market positions (Annual marker)

	Higher pay$_t$	*Low-pay$_t$*	*Total$_{t-1}$*
Higher pay$_{t\ 1}$	94.62	5.38	88.47
Low-pay$_{t-1}$	48.62	51.38	11.53
Total$_t$	89.32	10.68	100

Source: IDI data, own calculations. *N*=77 250.

Table A3: Regression results on low pay persistence (Annual marker)

	Low-pay$_t$
Labour market position at *t-1*	
Higher pay$_{t-1}$	*ref. category*
Low pay$_{t-1}$	1.159
	(0.044)
Labour market position at *t=0*	
Higher pay$_{t=0}$	*ref. category*
Low pay$_{t=0}$	2.121
	(0.063)
exogenous variables†	✓
σ_κ	1.484
	(0.036)
Log Likelihood	-18 630.421
N	77 250

Note: IDI (2018) and own calculations. Numbers in parenthesis refer to standard errors. Model contains same explanatory variables as the full model.

Table A4: Predicted probabilities of low pay persistence (Annual marker)

	Total	**Higher Pay$_{t=0}$**	**Low Pay$_{t=0}$**
$P(Higher\,pay_t \mid Higher\,pay_{t-1})$	0.951	0.977	0.800
	(0.078)	(0.017)	(0.111)
$P(Low\,pay_t \mid Higher\,pay_{t-1})$	0.049	0.023	0.200
	(0.078)	(0.017)	(0.111)
$P(Higher\,pay_t \mid Low\ pay_{t-1})$	0.880	0.932	0.582
	(0.144)	(0.044)	(0.150)
$P(Low\,pay_t \mid Low\ pay_{t-1})$	0.120	0.068	0.418
	(0.144)	(0.044)	(0.150)

Note: IDI (2018) and own calculations. Numbers in parentheses are standard errors.
N=77 250.

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.