ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH

Office for National Statistics

Using UK Tax Records to Produce New Statistics on Labour Market Transitions

Richard Dorsett and Jessica Hug

ESCoE Discussion Paper 2022-10

April 2022

ISSN 2515-4664

DISCUSSION PAPER

Using UK tax records to produce new statistics on labour market transitions
Richard Dorsett and Jessica Hug
ESCoE Discussion Paper No. 2022-10
April 2022

Abstract

This paper explores how we might use earnings data collected by HM Revenue Customs (HMRC) under the Pay As You Earn (PAYE) Real Time Information (RTI) system to construct novel statistics on individuals' employment and earnings transitions. Such data have great potential and offer particular strengths that are complementary to those of survey data. Nevertheless, they are collected for administrative rather than research reasons and their use in this latter regard is relatively recent. A consequence of this is that the PAYE RTI data variables available do not necessarily map directly onto the economic concepts of underlying interest. In this paper, we document the approach to constructing a dataset that can be used for the production of statistics on labour market transitions. We apply business rules intended to reduce the extent of spurious job transitions suggested by the raw data and to arrive at a consistent measure of weekly pay. With the resulting dataset we present a number of novel statistics. These are of interest in their own right but also serve to showcase the potential to make greater use of the PAYE RTI data and thereby gain new insights into the labour market.

Keywords: Administrative data; employment; earnings; transitions

JEL classification: J21, J31, Y10

Richard Dorsett, University of Westminster
r.dorsett@westminster.ac.uk

Published by:
Economic Statistics Centre of Excellence
National Institute of Economic and Social Research
2 Dean Trench St
London SW1P 3HE
United Kingdom
www.escoe.ac.uk

ESCoE Discussion Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the Economic Statistics Centre of Excellence (ESCoE), its partner institutions or the Office for National Statistics (ONS).

© Richard Dorsett and Jessica Hug

Using UK tax records to produce new statistics on labour market transitions*

Richard Dorsett† and Jessica Hug‡

March 2022

*This research has been funded by the Office for National Statistics (ONS) as part of the research programme of the Economic Statistics Centre of Excellence. The authors thank Francesca Foliano, Jamie Moore and Stefan Speckesser for their early contributions to the research. Helpful comments were received from Tom Evans, David Freeman, Debra Leaker and an anonymous reviewer. Any views expressed are solely those of the authors and so cannot be taken to represent those of the ESCoE, its partner institutions or the ONS. This work contains statistical data from HMRC which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information.

†University of Westminster

‡National Institute of Economic and Social Research

Abstract

This paper explores how we might use earnings data collected by HM Revenue Customs (HMRC) under the Pay As You Earn (PAYE) Real Time Information (RTI) system to construct novel statistics on individuals' employment and earnings transitions. Such data have great potential and offer particular strengths that are complementary to those of survey data. Nevertheless, they are collected for administrative rather than research reasons and their use in this latter regard is relatively recent. A consequence of this is that the PAYE RTI data variables available do not necessarily map directly onto the economic concepts of underlying interest. In this paper, we document the approach to constructing a dataset that can be used for the production of statistics on labour market transitions. We apply business rules intended to reduce the extent of spurious job transitions suggested by the raw data and to arrive at a consistent measure of weekly pay. With the resulting dataset we present a number of novel statistics. These are of interest in their own right but also serve to showcase the potential to make greater use of the PAYE RTI data and thereby gain new insights into the labour market.

Keywords: Administrative data; employment; earnings; transitions.
JEL Classification: J21, J31, Y10.

1 Introduction

The use of administrative data to produce labour market statistics is attractive for several reasons. Such data require no collection costs beyond those incurred for their administrative purpose, are typically available at scale and often at high frequency, and are not subject to selective non-response nor recall error among respondents. In this paper, we explore the potential to use UK tax records to construct novel statistics on individuals' payrolled employment and earnings transitions. These findings are of substantive interest in their own right but the primary purpose of the paper is to illustrate the potential of administrative data as a basis for producing labour market statistics and, in so doing, to argue the case for their greater use.

The Office for National Statistics (ONS) regularly publishes a joint report with HM Revenue and Customs (HMRC) providing labour market statistics derived from earnings data collected under the Pay As You Earn (PAYE) Real Time Information (RTI) system. These statistics are regarded as experimental, reflecting the fact that the methodologies used to produce them are still in their development phase. At the time of writing, the most recent release (January 2022) provides statistics on the following:

- Number of payrolled employees (levels and change (including inflows and outflows); monthly)
- Median and mean monthly pay (levels and monthly, quarterly, annual and biennial change; monthly)
- Pay distribution (10, 25, 50, 75 90 95, 99 percentiles)
- Variation by geography (local authority, NUTS1, NUTS2 & NUTS3), industry (SIC 2007 sectors), age group (under-18, 18-24, 25-34, 35-49, 50-64, 65+), NUTS1×sector and NUTS1×age.

These experimental statistics complement more established statistics based on surveys such as the Labour Force Survey (LFS), Annual Survey of Hours and Earnings (ASHE) and the Monthly Wages and Salaries Survey (MWSS). The latter have distinctive benefits; in particular, surveys provide flexibility with regard to the type of information that can be collected. The PAYE RTI data, for example, does not cover self-employed people or those out of work, nor does it provide detailed information on hours of work, occupation or the background characteristics of workers.[1] These relative strengths of survey data are sufficient to ensure the continued relevance of survey-based statistics.

Nevertheless, there is clearly scope to make fuller use of administrative data. Doing so capitalises on the the strengths of such data to allow new insights not possible using survey data. Moreover, where there are areas of overlap, an eventual possibility is that official statistics rely more on administrative data so that survey questionnaires can be shortened (reducing respondent burden and thereby perhaps increasing co-operation rates), or the available interview time can be used to collect information on alternative aspects of the labour market.

This paper focuses on the potential to use PAYE RTI data to produce novel statistics on labour market transitions. This is an area where administrative data has definite strengths,

[1] It also misses members of PAYE schemes where no employee earns above the Lower Earning Limit for National Insurance or has another job.

allowing near-costless tracking of individuals over an extended period of time. Survey data, by contrast, is compromised by non-response. As an illustration of this, in the final quarter of 2019 (before the onset of the COVID-19 pandemic) the proportion of eligible households responding at the fifth wave of the LFS was below one-third. While survey weights can go some way towards adjusting for the influence of such non-response, these do not address the possibility of unobserved influences on the response decision. Note also that the fifth wave of the LFS is the final interview for participating households and corresponds to a point roughly one year after the initial interview. Hence, aside from the non-response issue, transitions over more than 12 months are not possible using the LFS.

The remainder of the paper follows the following structure. In section 2, the data are described as well as an overview of the cleaning carried out prior to analysis. The statistics themselves are divided between three sections. In section 3, the focus is on payrolled employment transitions, both gross flows into and out of payrolled employment. In section 4, attention turns to jobs rather than payrolled employment. Lastly, in section 5, we provide evidence on earnings transitions. Section 6 briefly concludes.

2 The PAYE RTI data

Since April 2014, all employers have been required to send information about tax and other deductions under the PAYE system to HMRC every time an employee is paid. The analysis in this paper is based on these data and covers the population of employee jobs for the tax years 2014/15, 2015/16, 2016/17 and 2017/18.

The raw data were subject to initial processing by HMRC and aggregated to be on a monthly basis. Each record in the resulting data relates to a single job and includes unique individual and employer identifiers. The data require cleaning before they can be used for analytical purposes. Full details of the business rules applied to achieve this are provided in Appendix A. In brief, they aim to correct for likely reporting irregularities that erroneously suggest breaks in job spells or negative earnings in some months. It is not always appropriate to interpret recorded earnings in a month as corresponding to employment in that month, nor the absence of recorded earnings as non-employment. A disparity would arise, for instance, where an employer was late in submitting a return to HMRC.

Such recording issues are relevant for statistics on labour market transitions since they can create a misleading impression of job, employment or earnings mobility. A key aim of the data processing is to impute payrolled employment from earnings in a sensible way, smoothing spells when appropriate but not losing true breaks in jobs.

The key variable for our analysis is monthly earnings. This is constructed as taxable pay plus contributions to occupational pension schemes, payroll giving (charitable donations) and the value of childcare and other non-cash vouchers which are assessed for National Insurance Contributions. Benefits in kind are excluded. This definition is intended to represent as closely as possible the measure of headline gross pay that employees see on their pay slips (and so is consistent with a common understanding of total pay). After appropriate cleaning has been conducted, a job is assumed to exist in a given month if it has positive earnings within that month. Reading across all jobs for an individual gives a measure of payrolled employment that

allows for multiple job-holding.

3 Payrolled employment transitions

The results presented in this section are at the level of the employee rather than the job. As with the results in section 4, the statistics highlight two strengths of the data: long-term tracking (individual transitions over a two-year period) and large size (allowing highly disaggregated analysis, showcased here through statistics showing variation at the level of the travel-to-work area (TTWA)). We emphasise that, since self-employment is not observed, the reported transitions are more accurately described as movements into and out of *payrolled* employment. We maintain this terminology throughout.

3.1 Gross-flows out of payrolled employment

With longitudinal monthly data, we are able to produce statistics on both short-term and longer-term transitions. Figure 1 shows the proportion of April 2015 employees who are out of payrolled work 1, 2, 12 and 24 months later.[2] This ranges from 2.5% within 1 months to 15% within 2 years. We note that this does not necessarily imply that they are not working; they may instead be self-employed. There is limited information on the background characteristics of employees. However, gender and age are recorded. The pattern of transitions does not vary substantially by gender. By contrast, there is a clear U-shaped age profile, with younger and older individuals more likely to be out of payrolled work than prime-age workers. This is true across all measures. There are a number of familiar potential explanations. For young workers, jobs may be more precarious until they have acquired sufficient experience to be valuable. Alternatively, they may still be in the process of trying out jobs – 'job shopping' – or may decide to return to full-time education. For older workers, the higher rates of being out of payrolled employment may reflect difficulties finding work should they lose their initial jobs or, particularly for the oldest group, retirement decisions.

Figure 2 shows variation by earnings decile. With the exception of the highest decile, the probability of being out of payrolled work decreases with earnings. This is true across all durations. The proportion out of work 24 months later is roughly 3 times higher among those in the first decile compared to the ninth decile. For shorter-term outcomes, the proportion out of payrolled work is roughly 10 times higher for the bottom decile than it is for the ninth decile.

We can also see how payrolled employment transitions vary according to the characteristics of jobs.[3] Figure 3 shows wide variation by industry in the proportion of April 2015 employees out of payrolled work 1, 3, 12 or 24 months later.[4] This ranges from 1% in Electricity, Gas, Steam and Air Conditioning Supply to 4% in Administrative and Support Service Activities after 1 months and from 9.5% in Public administration to 22% in Activities of Households as Employers after 24 months.

[2]The results are comparable when based on the population of employees in April 2014 or April 2016.

[3]This is information originates from the Inter-Departmental Business Register (IDBR), to which the PAYE RTI data were linked

[4]Where a group of enterprises is under common ownership, industry is defined at the group level.

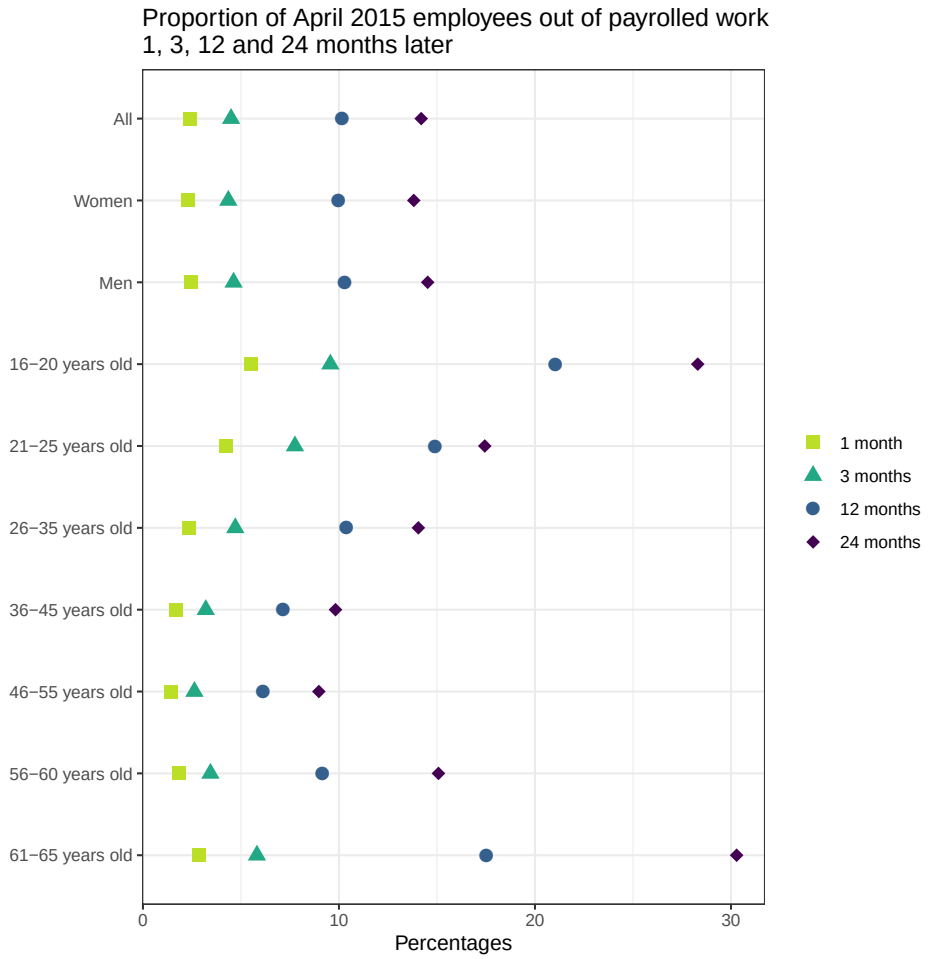


Figure 1
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by age take the situation of the worker in April 2015. 'Out of work' is defined here as the absence of regular payrolled work.

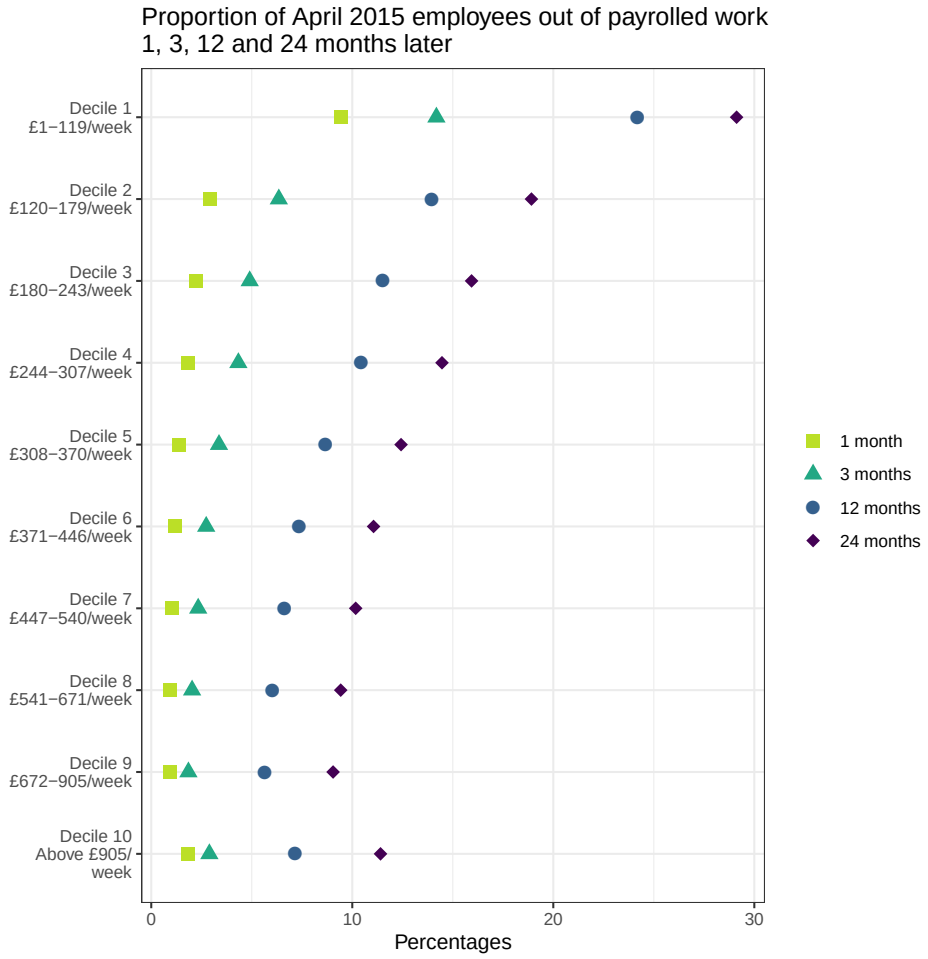


Figure 2
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: 'Out of work' is defined as the absence of regular payrolled work. Deciles of UK pay in April 2015.

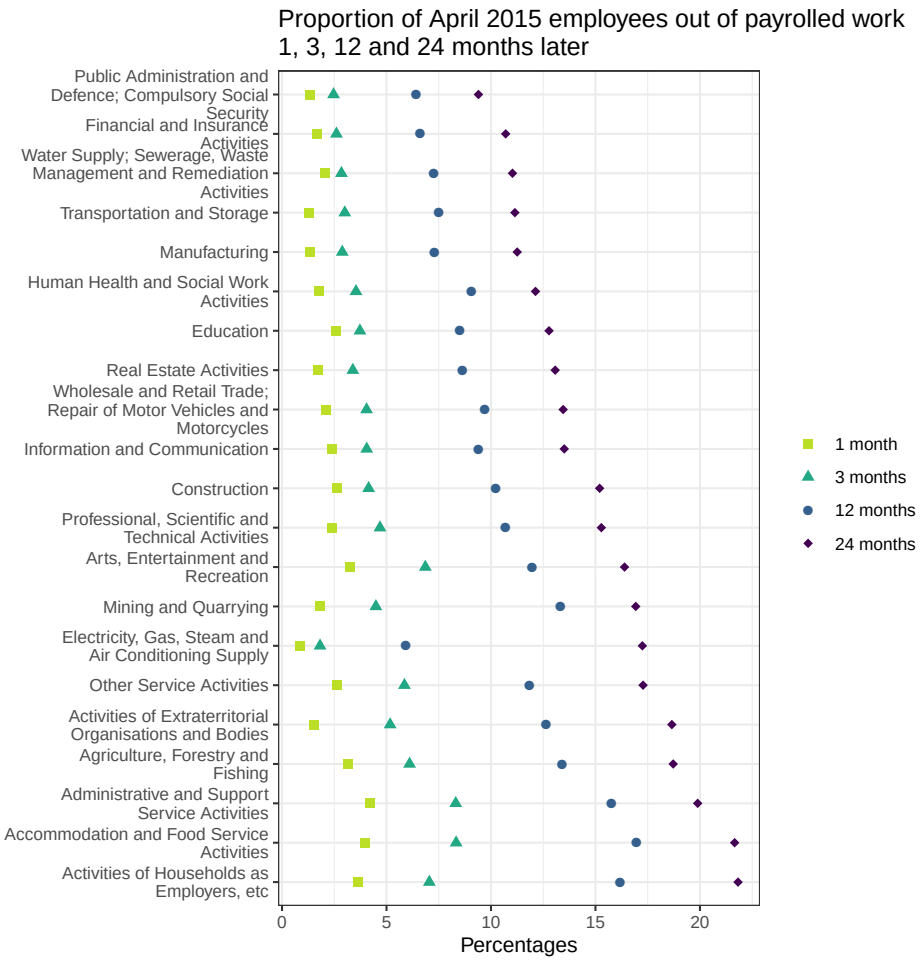


Figure 3

Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.

Notes: Industry relates to April 2015. Where an individual has multiple jobs, each job is included in the relevant industry, weighted by the inverse of the number of jobs held by the individual at that time (so the contribution of the individual remains as 1). 'Out of work' is defined as the absence of regular payrolled work.

3.1.1 Taking advantage of the large sample size: small area analysis

Figure 4 shows variation by region of residence. The main impression is of London looking distinct from other regions.

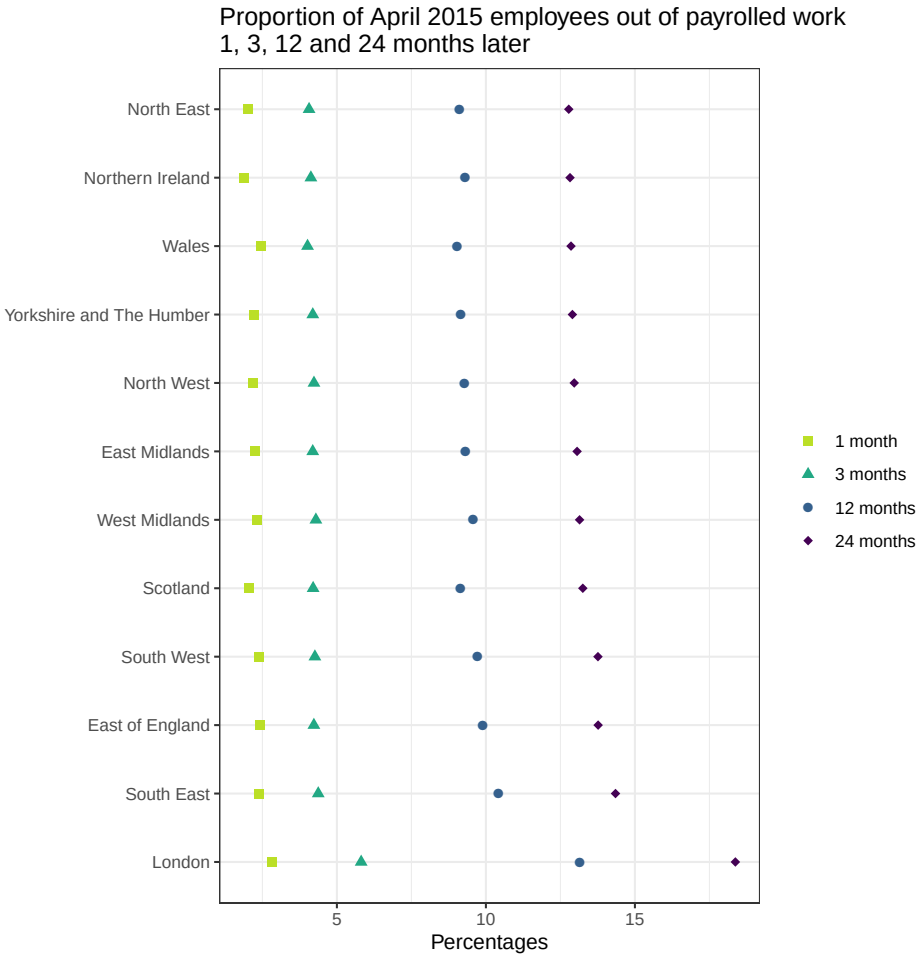


Figure 4
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in April 2015. 'Out of work' is defined as the absence of regular payrolled work.

The large number of observations available allows us to produce robust statistics at a finer level of geographical detail. Figure 5 suggests considerable heterogeneity across TTWAs in payrolled employment retention.[5] While Figure 4 might suggest little geographic variation, the ability to focus on smaller areas reveals a more detailed pattern. Unlike the statistics presented above, Figure 5 shows results for two populations: April 2014 employees and April 2015 payrolled employees. This allows the stability of the geographical pattern to be assessed.

For the April 2015 employee population, the proportion out of payrolled work one month later ranges between 1.5 and 4%. At 24 months, the corresponding range is from 10 to 19%. A comparison with the April 2014 employee population shows differences for some travel to work areas. This may capture genuine economic differences, such as different trends or reactions over

[5] As with region, this is based on residence rather than workplace.

the business cycle, but it may also reflect an increased random variation due to TTWA-level statistics being based on a smaller population. However, the fact that it is not among the smallest-population TTWAs that the most variation is seen between 2014 and 2015 perhaps suggests that the differences do tend to reflect underlying economic changes (local shocks such as the closure of big plants).

However, there is also a considerable degree of consistency across 2014 and 2015. This is particularly true in the longer term; the variation across TTWAs in the proportion out of payrolled work after two years is nearly identical across the two populations. The within-region variation is often quite marked (Scotland is perhaps most notable in this regard).

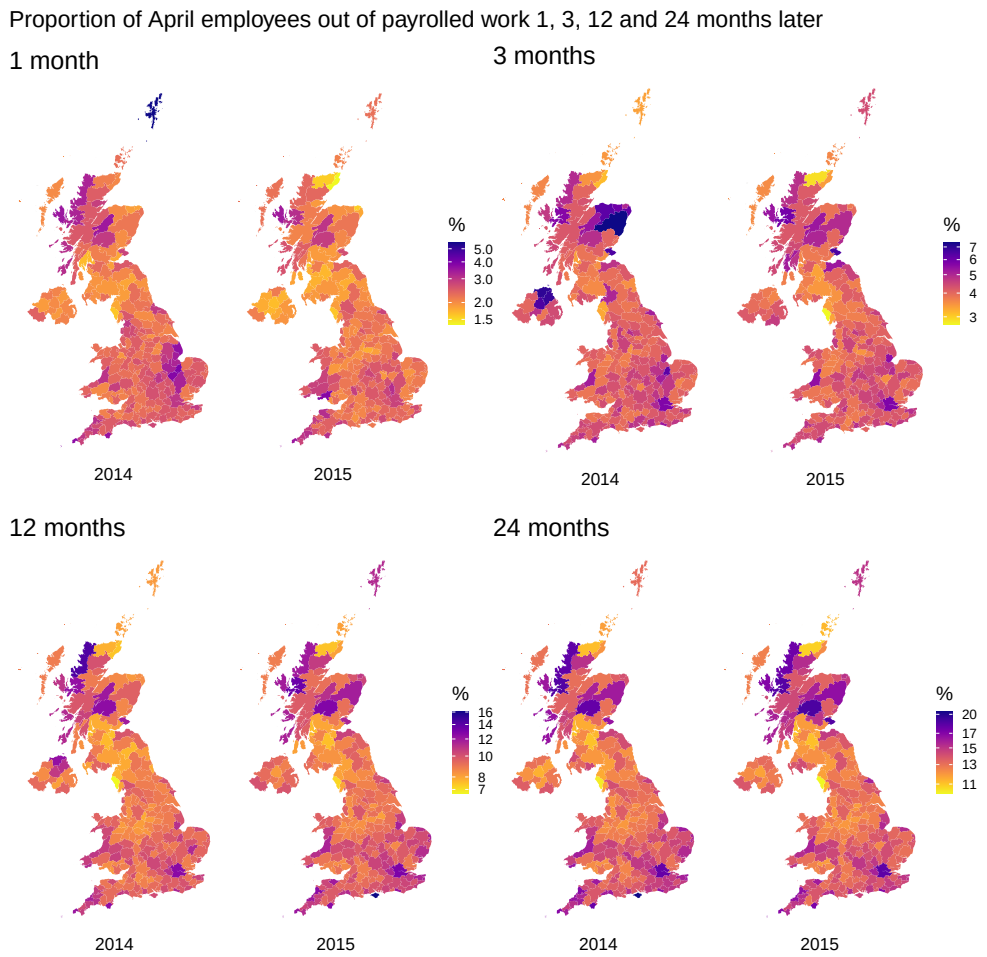


Figure 5
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in April each year. 'Out of work' is defined as the absence of regular payrolled work.

As a demonstration of this consistency, Figure 6 graphs the 2014-2015 comparison in TTWA proportions out of payrolled work. In line with the visual impression from 5, these charts show that the correlation becomes stronger the longer the outcome period considered. This perhaps reflects the tendency for short-term comparisons to be more influenced by local shocks that even out over the longer term.

The scale of the PAYE RTI data is such that it even possible to look at subgroups by TTWA.

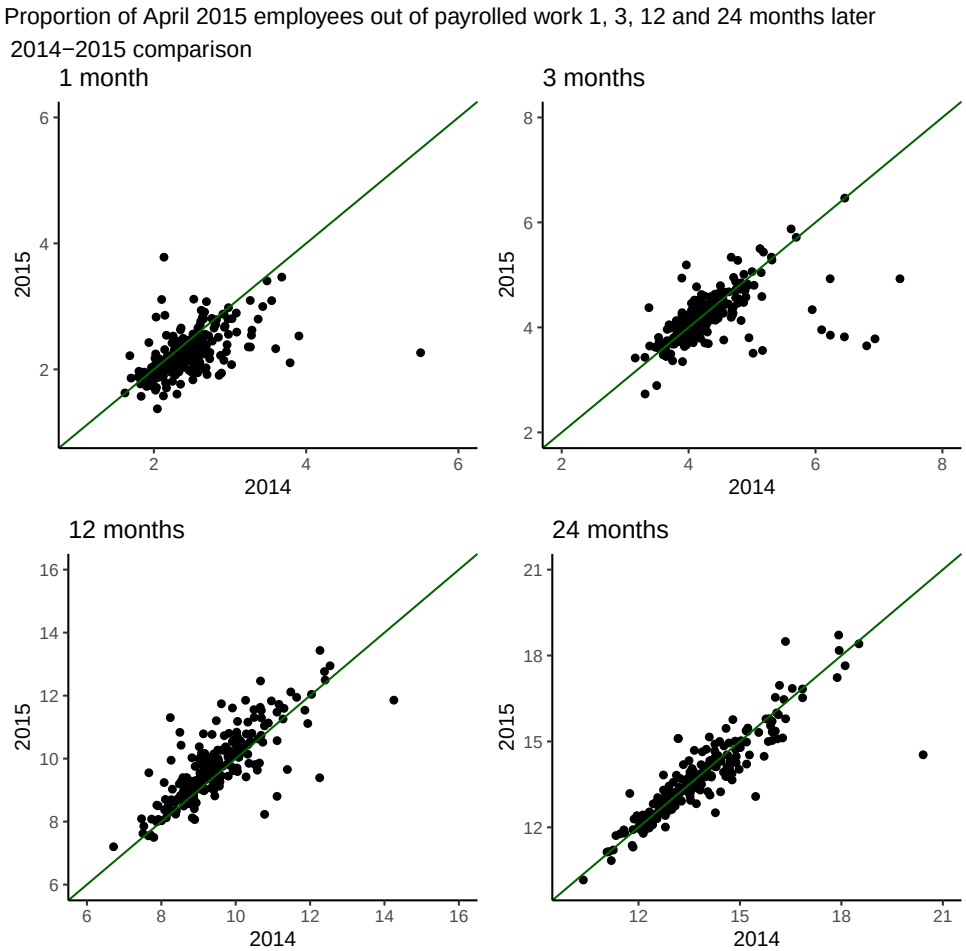


Figure 6
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in April each year. 'Out of work' is defined as the absence of regular payrolled work.

Figure 7 shows the proportion of April 2015 employees out of payrolled work 24 months later, broken down by a range of characteristics. It is with earnings level that the starkest differences are seen; those earning below the median are much more likely to be out of payrolled work two years later than those earning above the median. Nevertheless, there is considerable variation across TTWAs within most sub-groups. For instance, among employees earning above the UK median, the probability of being out of payrolled work ranges from 6 to 18% while for employees aged 21-24 the range is 9.5 to 25%. London has one of the highest proportion of employees out of payrolled work in all subgroups except employees aged above 55.

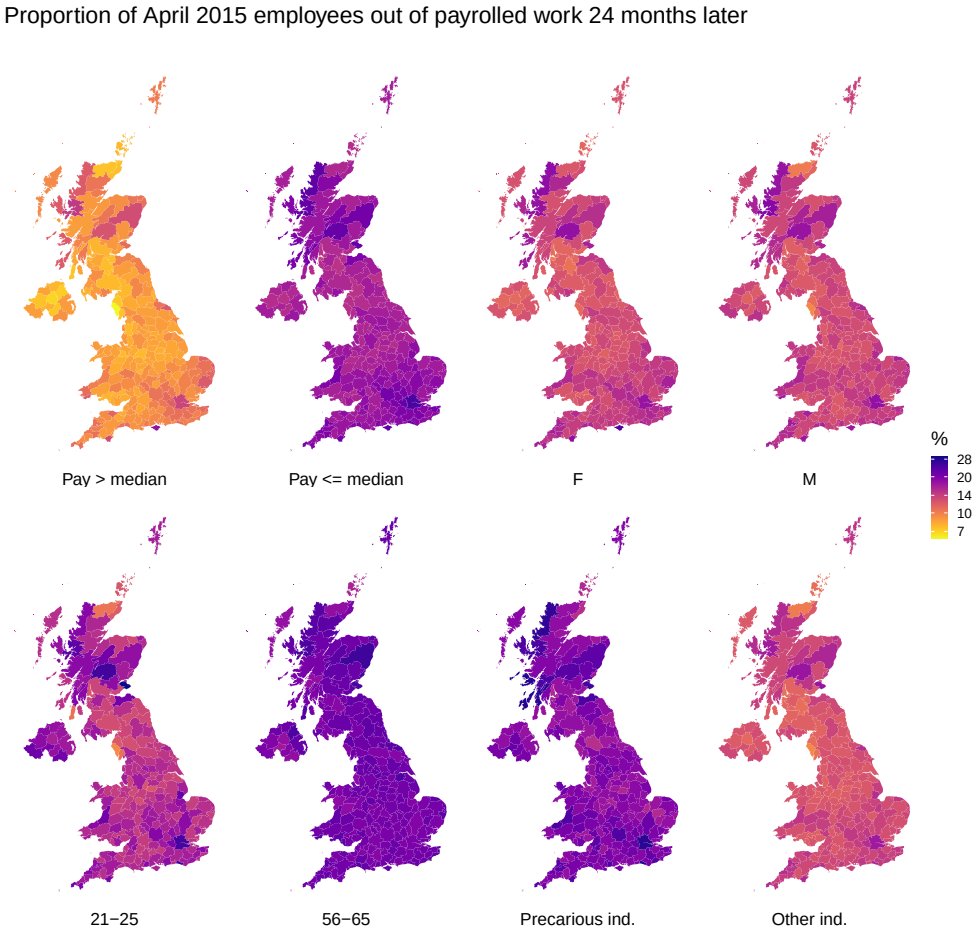


Figure 7
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by age, geographies and industries take the situation of the worker in April 2015. 'Out of work' is defined as the absence of regular payrolled work. Precarious industries are Accommodation and Food Service Activities as well as Activities of Households as Employers. Note that the maps with pay below and above the median refer to 2014.

3.2 Gross-flows back into payrolled employment

In this sub-section, attention turns from the employed population to the non-employed or, more accurately, those out of payrolled employment. Doing so highlights a fundamental characteristic of the PAYE RTI data; while they capture the full population of payrolled employees, they do not provide any information on those not in payrolled employment. This is an important limitation and means that it is not possible to consider flows between out-of-payrolled-work states.

However, for those who are employees at some point, the longitudinal nature of the data allows us to observe when a worker ceases to be payrolled. We make use of this to present statistics on the those individuals who were in payrolled work at any point in 2014 but out of payrolled work in April 2015. This sample is not representative of the April 2015 out of work population since, in addition to point already made that some of those no longer in payrolled employment may be self-employed, it excludes unemployed and economically inactive people who were not working at some point in 2014. Nevertheless, it is informative of the payrolled employment re-entry chances of the 2014 employee population and as such can provide insight into the degree of labour market resilience and how this varies with individual and area characteristics. Do some groups who leave payrolled employment find it more difficult to get a new job? Are some areas better equipped to re-employ displaced workers?

Figure 8 shows, for those not in payrolled employment in April 2015 but who were at some point during the preceding year, the proportion back in payrolled work 1, 3, 12 and 24 months later. Contrary to gross flows out of payrolled employment, men and women differ in this regard. For instance, 33% of men were back in payrolled work 24 months later compared to 37% of women. The 21-25 age group has the highest proportion of individuals back in payrolled work after 25 months (46%), much higher than the 26-30 age group (35%), perhaps reflecting an enduring cohort penalty of entering the labour market at the time of the financial crisis. Another explanation may be graduates moving into payrolled employment (provided they were payrolled employees at some point in the previous year.) The low proportion back in payrolled work among the 16-20 age group might be explained by returning to full-time study, while the low proportions among older workers may partly be due to early retirement.

Figure 9 shows less variation across earnings deciles than for the proportions of workers leaving payrolled employment. The second decile has the lowest proportions back into payrolled work (33.5% 24 months later), the first decile does slightly better (34.5% 24 months later). The differences are small though. Again, this top decile looks somewhat distinct from the immediately preceding deciles.

Figure 10 shows the breakdown by broad industry category of the new employer. Those entering work concentrate in a relatively small number of industries. This is particularly apparent at longer durations since April 2015.

3.2.1 Taking advantage of the large sample size: small area analysis

Figure 11 shows that, as with the probability of being out of payrolled work, the probability of payrolled employment re-entry is fairly stable across regions. Again, there is a distinct London effect. There are also slightly higher probabilities in Northern Ireland.

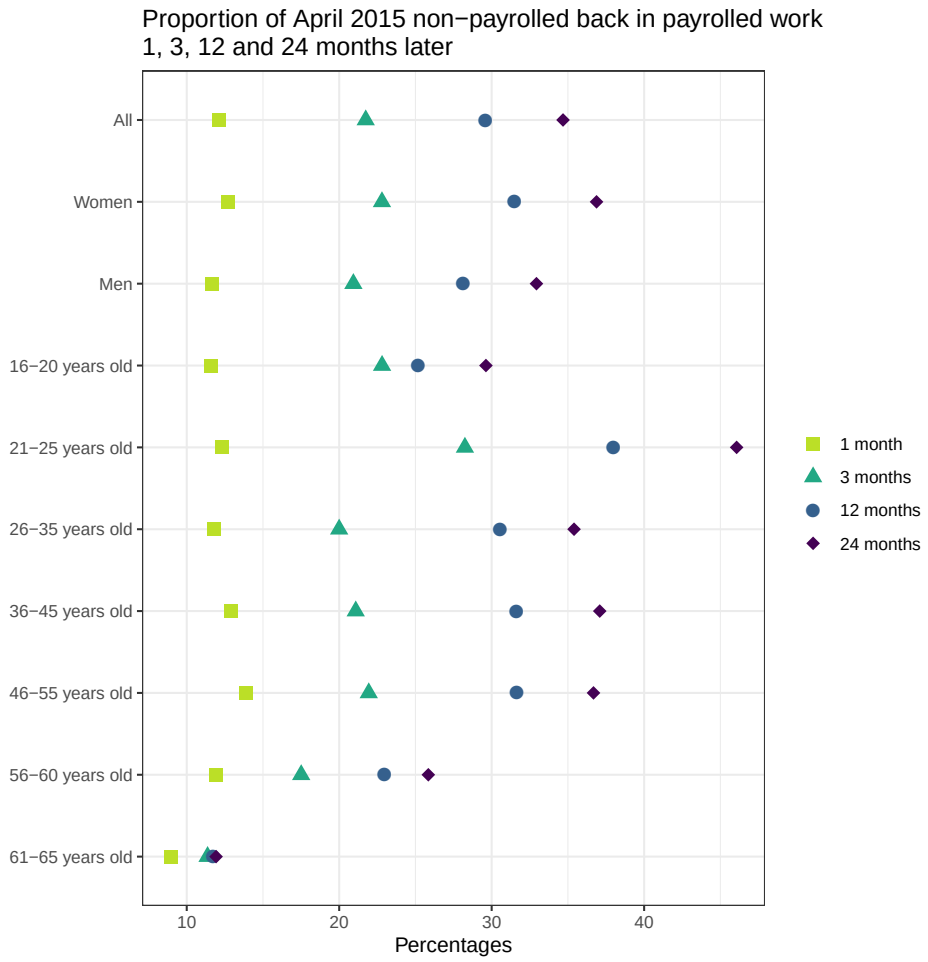


Figure 8
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by age take the situation of the worker in April 2015. 'Out of work' is defined as the absence of regular payrolled work. The base for computing percentages comprises all individuals who were non-payrolled in April but payrolled at some point in the preceding year.

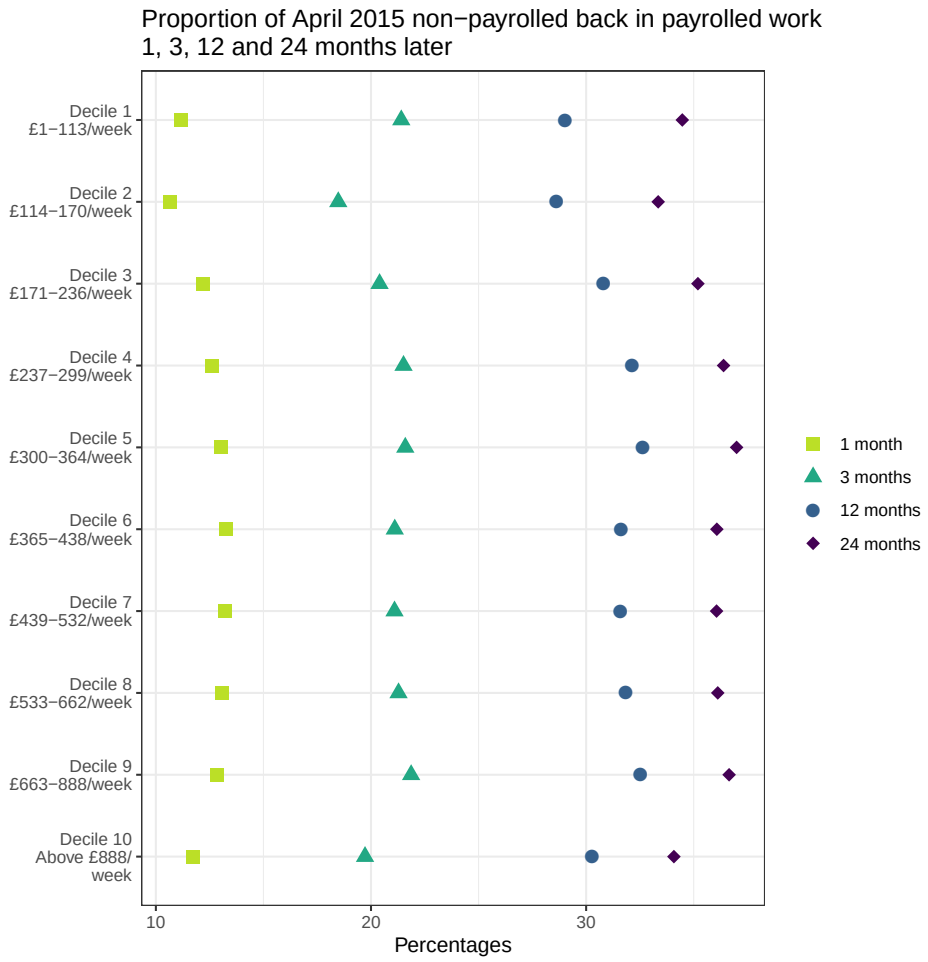


Figure 9

Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.

Notes: the distribution of earnings is from workers in payrolled employment in April 2014. The base for computing percentages comprises all individuals who were non-payrolled in April but payrolled at some point in the preceding year.

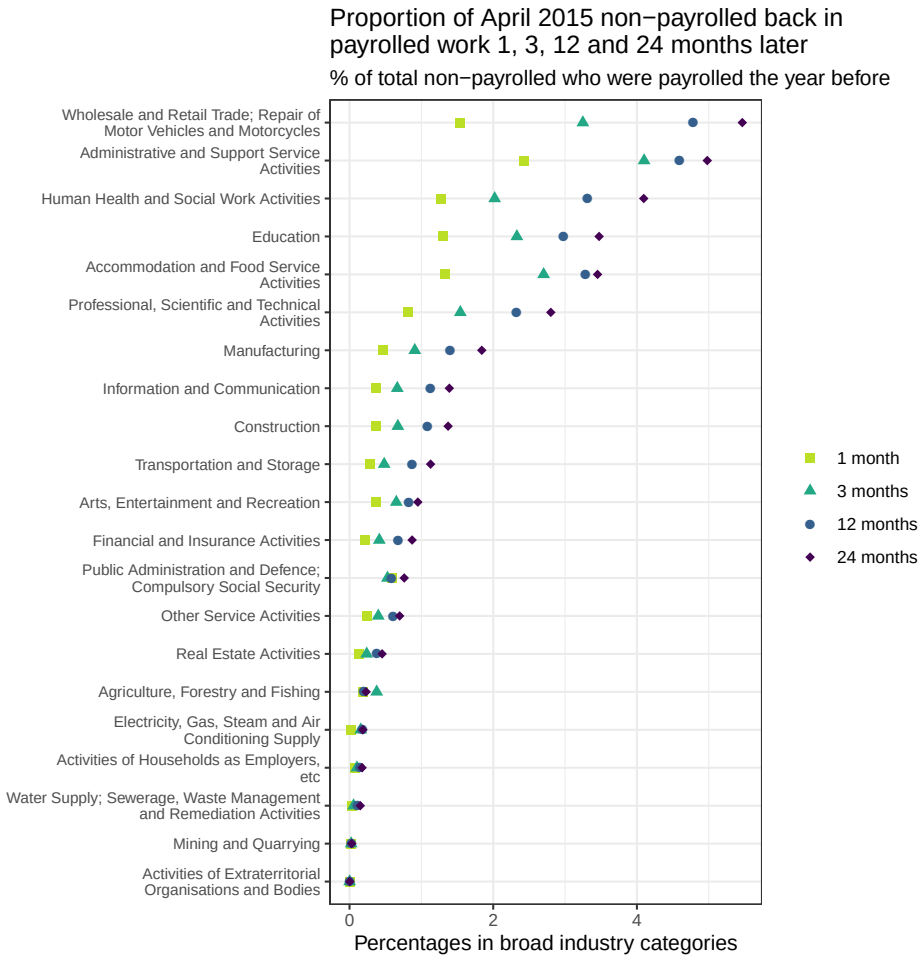


Figure 10

Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.

Notes: The base for computing percentages comprises all individuals who were non-payrolled in April but payrolled at some point in the preceding year.

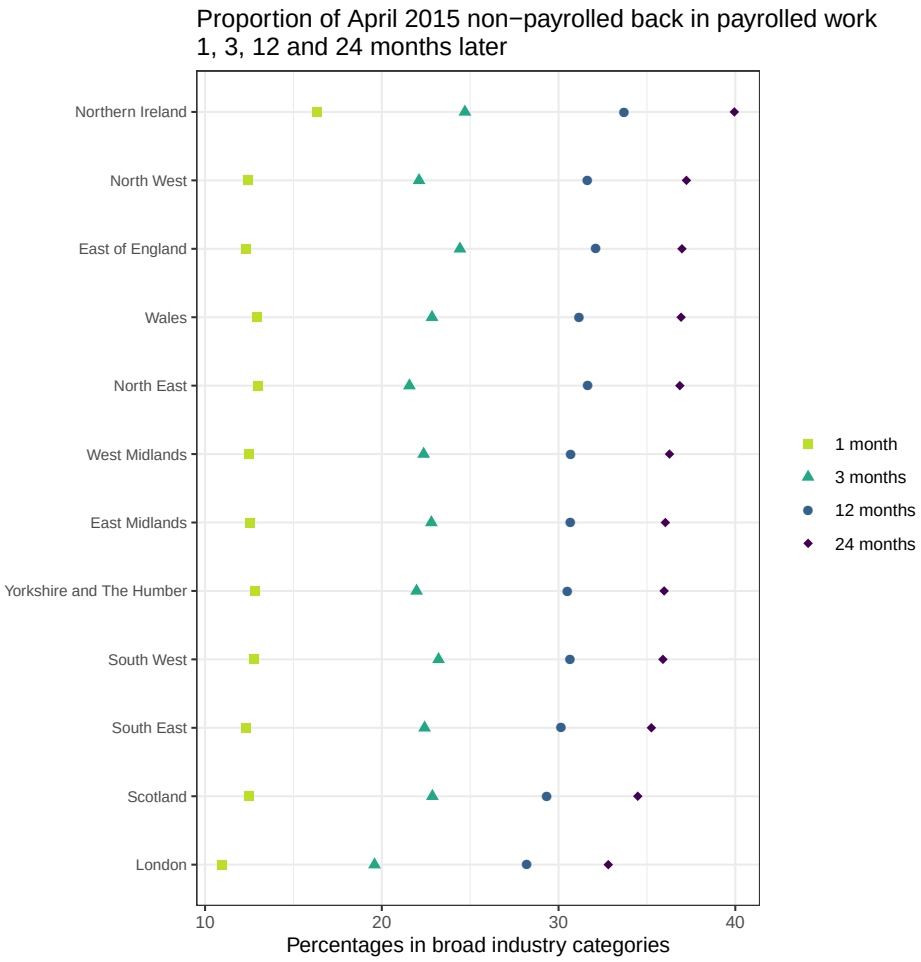


Figure 11
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in April 2015. The base for computing percentages comprises all individuals who were non-payrolled in April but payrolled at some point in the preceding year.

As before, we are able to inspect variation by TTWA. Figure 12 reveals differences across the country but also within region. With longer-term outcomes, the broad impression is of a North-South divide but with a number of local black spots, particularly in coastal areas.

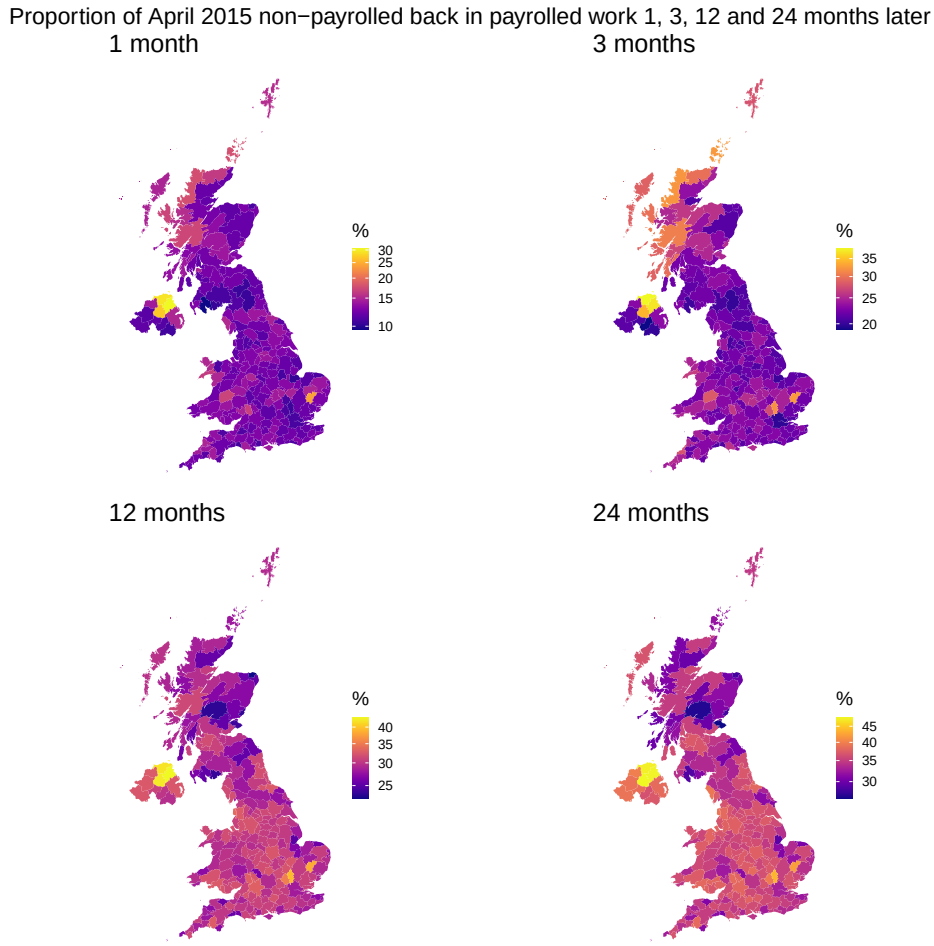


Figure 12
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in each year.

Figure 13 provides an indication of how this variation differs by subgroup. Unlike the analysis of payrolled employment exits, the probability of being back in payrolled work does not appear to vary so much with the level of pay when previously working. However, there is perhaps more TTWA variation among the higher paid than the lower paid.

4 Job transitions

The results presented in this section are at the level of the job rather than the individual. This follows a similar format to section 3 but allows variation by employer characteristic as well as employee characteristic to be explored. This demonstrates another advantage of the data;

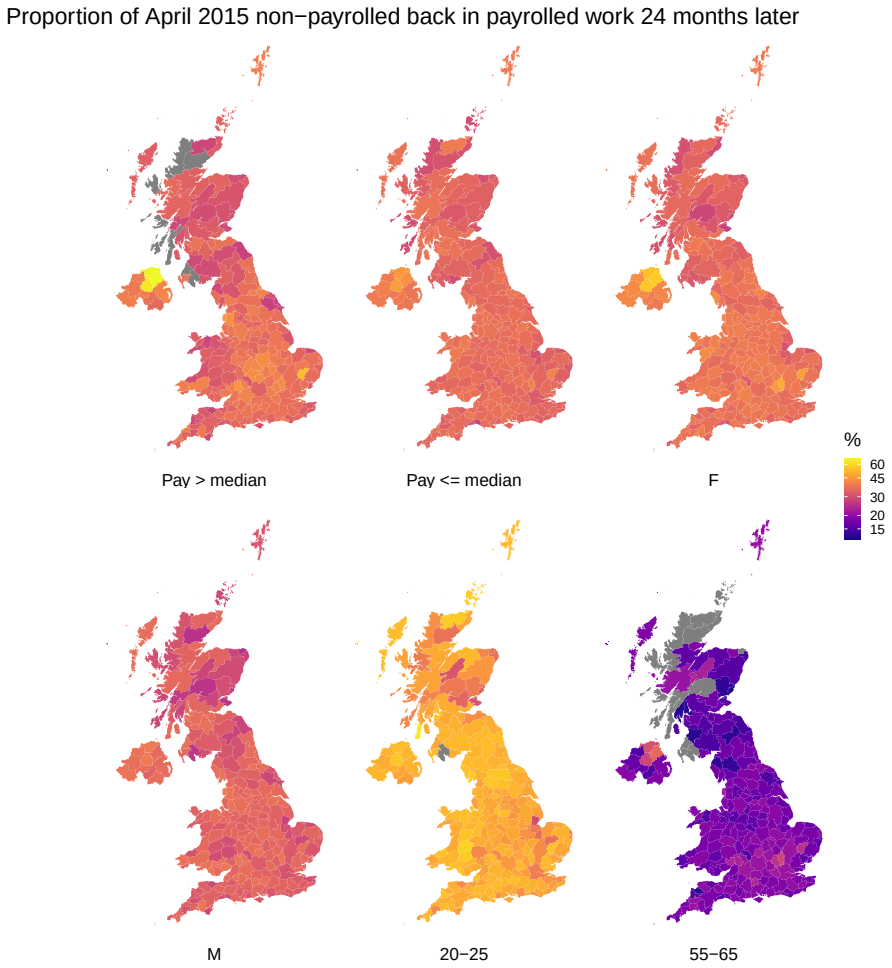


Figure 13
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by age, geographies and industries take the situation of the worker in April 2015. Grey TTWAs are TTWAs with suppressed statistics as they are based on a count of less than 30 observations. Re-entering by industries' group is not displayed as the counts does not allow it contrary to Figure 7.

since all payrolled employees are observed, it is possible to construct descriptive measures of employers, such as their size.

4.1 Job duration

Figure 14 shows the proportion of April 2015 jobs that are ongoing 1, 3, 12 and 24 months later. More than 60 per cent of jobs are still live two years later and there is no difference between men and women. Age is a strong predictor of job duration. Among under 25 year-olds only about 40% of jobs last two years, compared to 73% for 46-55 years old.

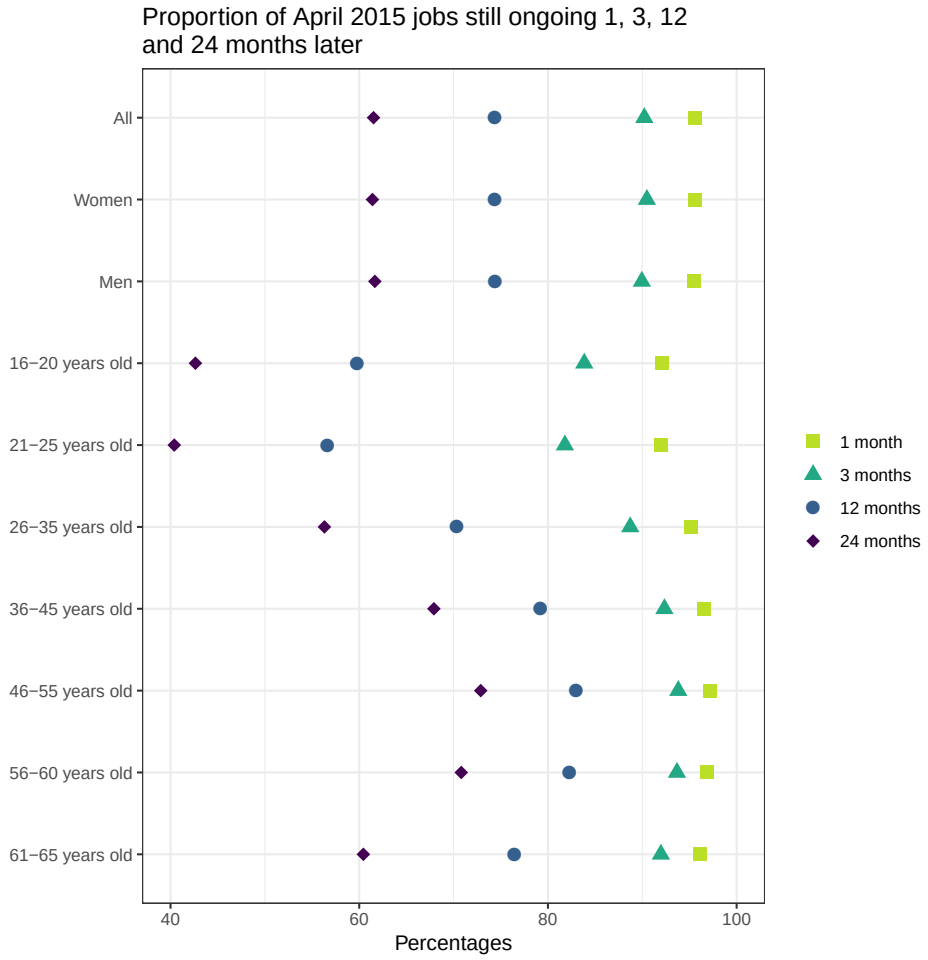


Figure 14
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: workers can have many jobs. Different jobs by the same employer are regarded as a single job. The breakdowns by age take the situation of the worker in April 2015.

Figure 15 shows that there is a high degree of variation across industries in the longevity of jobs. Public administration is the most stable industry with 80% of jobs still ongoing 24 months later. At the other extreme are accommodation and food service activities with only 39% of jobs ongoing 24 months later.

Figure 16 shows the relationship with firm size.[6] This is U-shaped; middle-sized firms (those

[6]Where a group of enterprises is under common ownership, firm size is calculated at the group level.

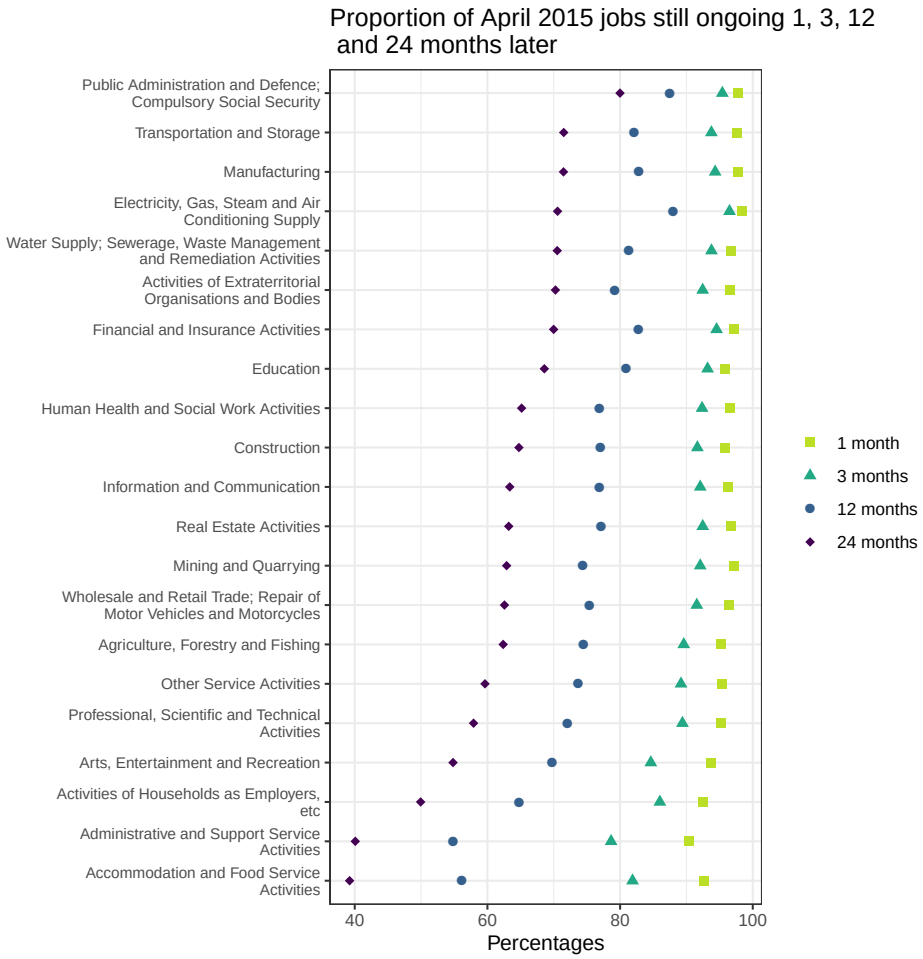


Figure 15

Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.

Notes: workers can have many jobs. Different jobs by the same employer are regarded as a single job. Broad industry categories according to the 2017 SIC definition.

with 51 to 100 employees) have the shortest job durations, with 57% of jobs ongoing 24 months later compared to 63 and 64% among the largest and smallest categories, respectively.

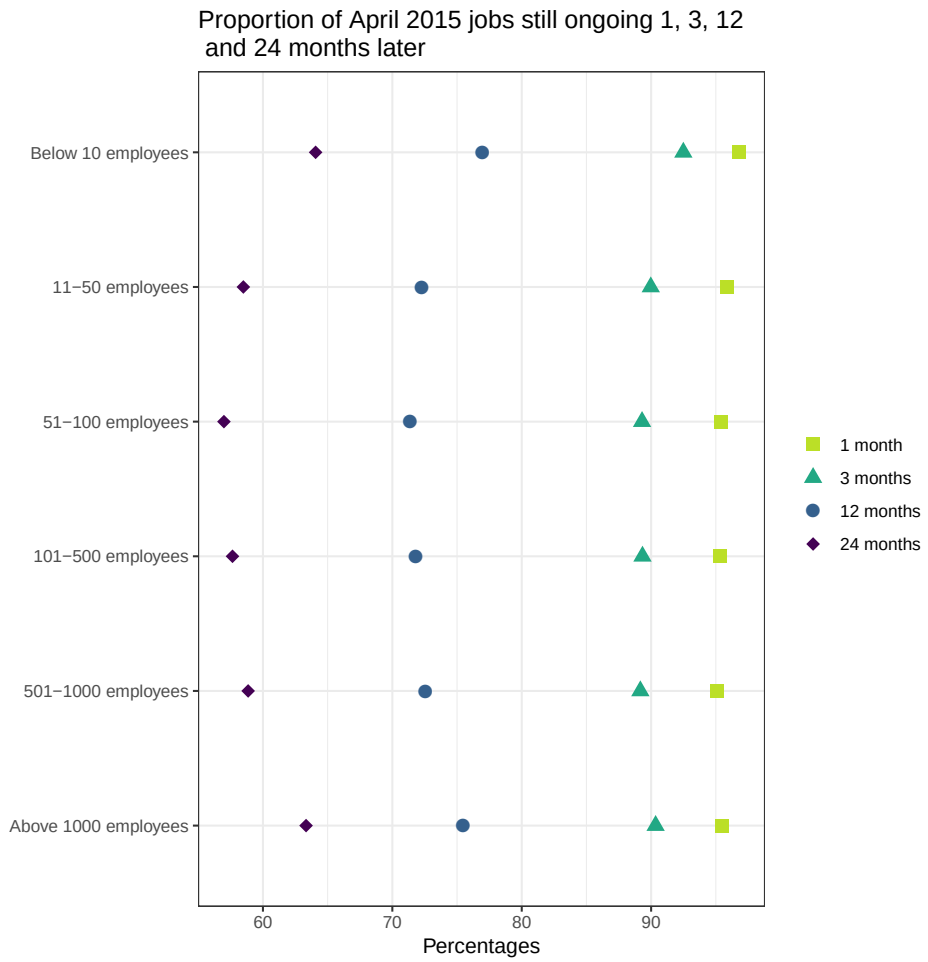


Figure 16
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: workers can have many jobs. Different jobs by the same employer are regarded as a single job.

4.1.1 Taking advantage of the large sample size: small area analysis

Figure 17 suggests regional variation in job duration. Longer job duration is seen in Northern Ireland, Scotland and Wales. In Northern Ireland, 67.5% of jobs were ongoing 24 months later, compared to 61% in the South East and South West of England. As seen before, London is again distinctive with only 55% of jobs lasting two years.

A more complex pattern is evident in Figure 18. As with flows out of payrolled employment, TTWAs in Scotland are very different in terms of job duration, where TTWAs with the highest flows out of payrolled employment are also the ones with the shortest job duration. The visual impression when comparing April 2015 jobs with April 2014 jobs is one of broad consistency in the pattern of TTWA variation. This is perhaps particularly the case when considering longer-term outcomes. While this was also the case when considering payrolled employment (Figure 7) rather than job duration, the inference differs. Whereas a higher probability of being out

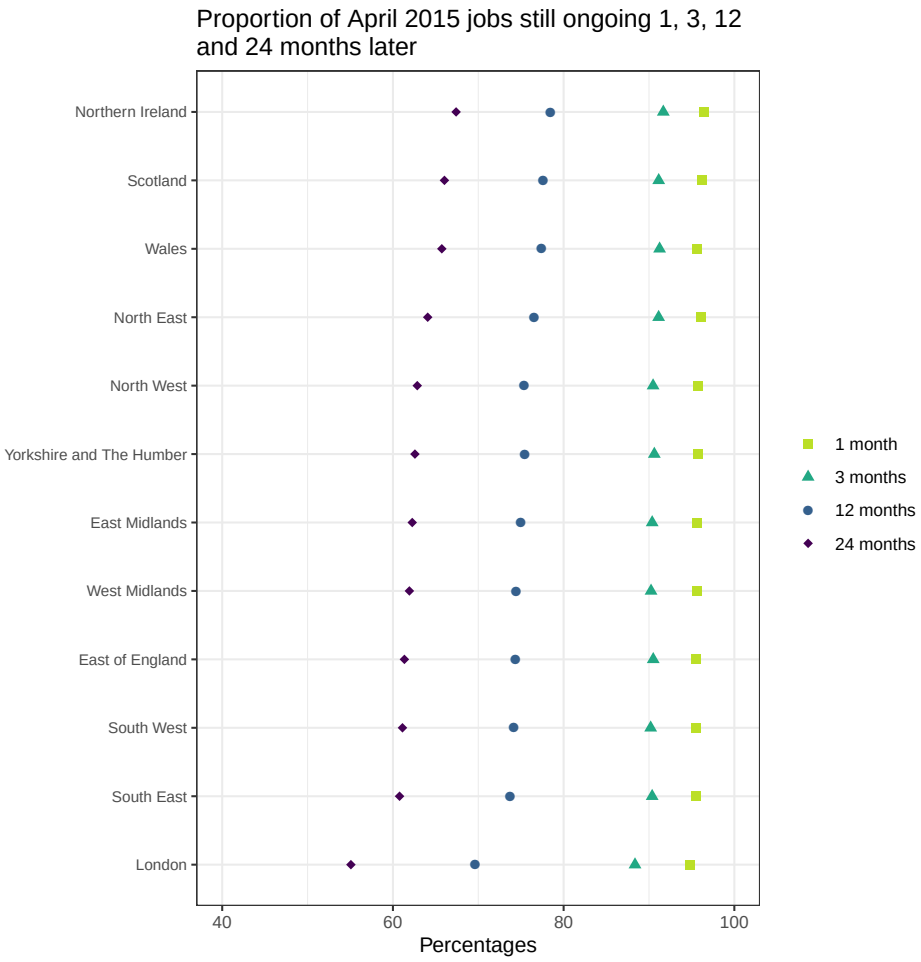


Figure 17
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: workers can have many jobs. Different jobs by the same employer are regarded as a single job. Broad industry categories according to the 2017 SIC definition. The breakdowns by geographies take the situation of the worker in April 2015.

of payrolled work is likely to reflect a weaker local labour market, shorter job duration is also consistent with a dynamic local labour market that may offer ready employment opportunities.

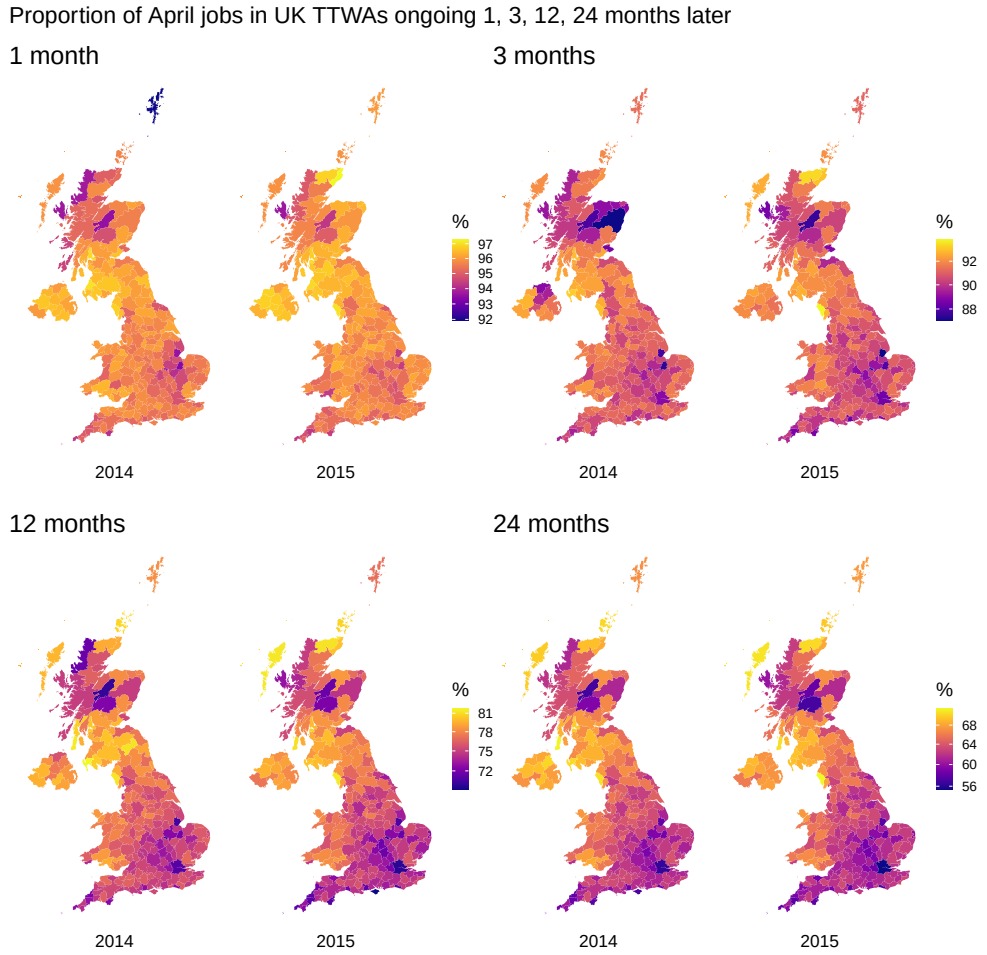


Figure 18
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: workers can have many jobs. Different jobs by the same employer are regarded as a single job. Broad industry categories according to the 2017 SIC definition. The breakdowns by geographies take the situation of the worker in April 2015.

4.2 Job to job transitions

The data allow us to examine job-to-job transitions. This is an analysis that operates at the level of the individual but requires information at the level of the job. Since the PAYE RTI data records all employee jobs, it is ideal for probing more nuanced aspects of job transitions. We highlight the frequency of three types of transition among payrolled employees: a change from holding a single job to holding a different single job; a change in the composition but not number of jobs for those holding multiple jobs; and an increase in the number of jobs held.

Figure 19 considers the case of single job-holders changing job. It shows that 20% of all payrolled employees will have changed jobs within two years. This is similar for men and

women but there is a strong age profile, with young people more likely to change; 35% of 21-25 year-olds changed jobs within 24 months, compared to less than 25% for all older age categories.

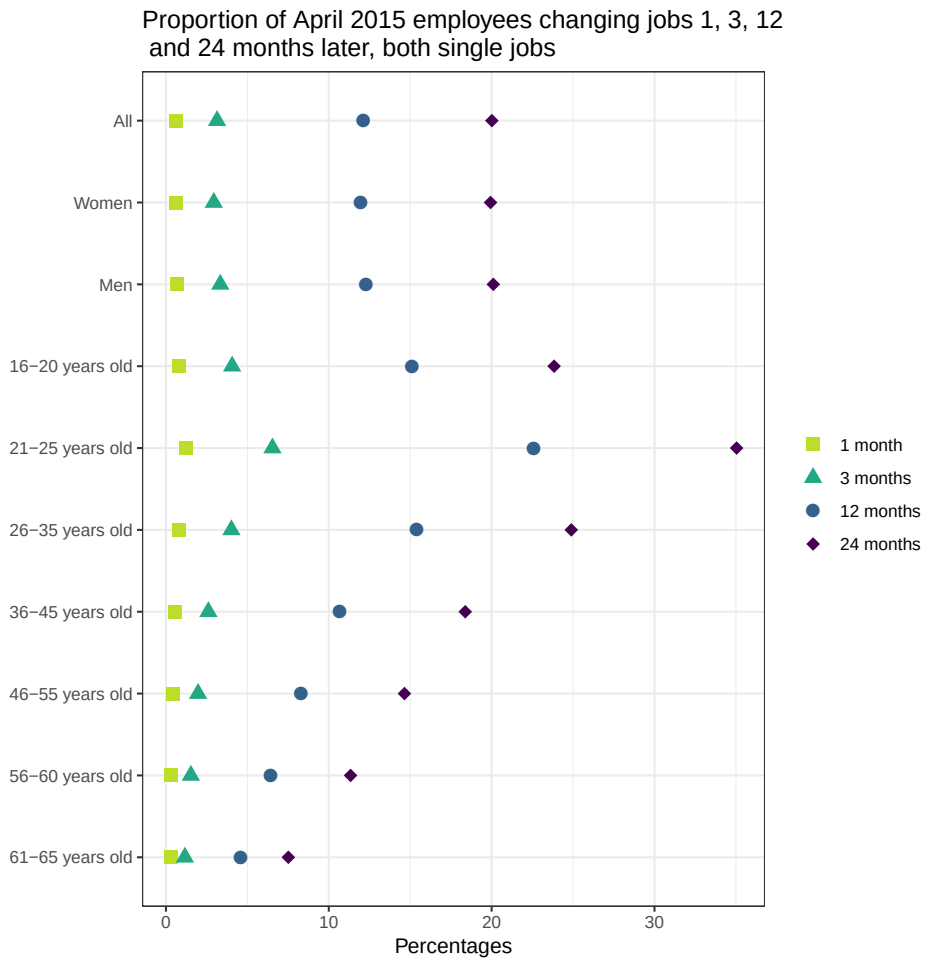


Figure 19
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by age take the situation of the worker in April 2015. Proportion of transitions in a different broad industry: 1 month: 60%, 3 months: 62%, 12 months: 61%, 24 months: 82%.

Figure 20 shows the proportion of all payrolled employees who change one or more jobs when holding multiple jobs, while keeping the overall number held stable. Here, there is a gender difference, with women more likely to change jobs than men. This is either because women tend to hold more second jobs or because women who have multiple jobs change jobs more than men do. As with individuals holding single jobs, the highest transition rate is for the 21-25 age category with 65% changing at least 1 job. In other regards, the age profile is different from that for single job holders, and does not show a clear pattern.

Figure 21 considers the proportion of workers who take on additional jobs. Women and younger people are most likely to do this. Among female workers, 3.2% take up an additional job within 24 months, compared to 2.2% for men. Among employees aged 21-25, 4% take up an additional job within 24 months; that proportion falls among older age groups.

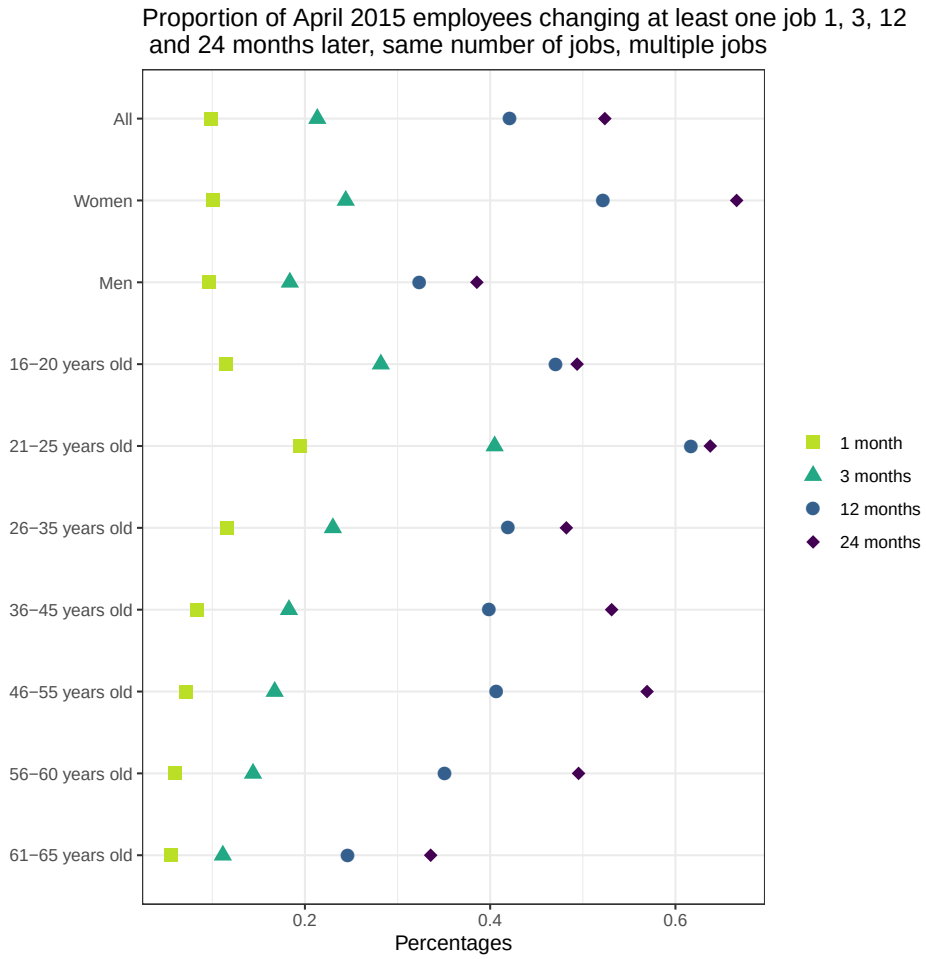


Figure 20

Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.

Notes: the breakdowns by age take the situation of the worker in April 2015. Proportion of transitions in a different broad industry: 1 month: 62%, 3 months: 63%, 12 months: 82%, 24 months: 82%.

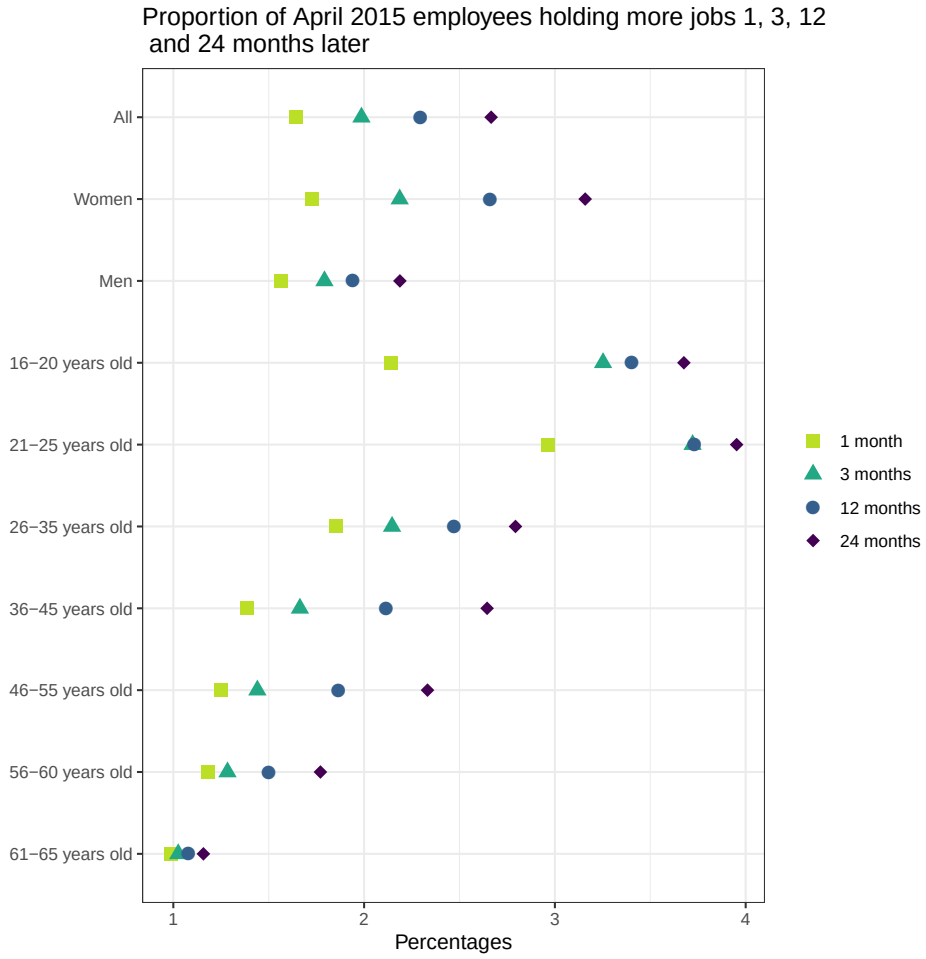


Figure 21
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by age take the situation of the worker in April 2015.

4.2.1 Taking advantage of the large sample size: small area analysis

Figure 22 shows the variation across TTWAs in the proportion of single job holders who change to a different (single) job within 1, 3, 12 and 24 months. It is apparent, particularly as longer-term outcomes are considered, that England is quite distinct from the other countries in the UK, and shows a higher rate of change in job. Substantial within-region variation is also evident.

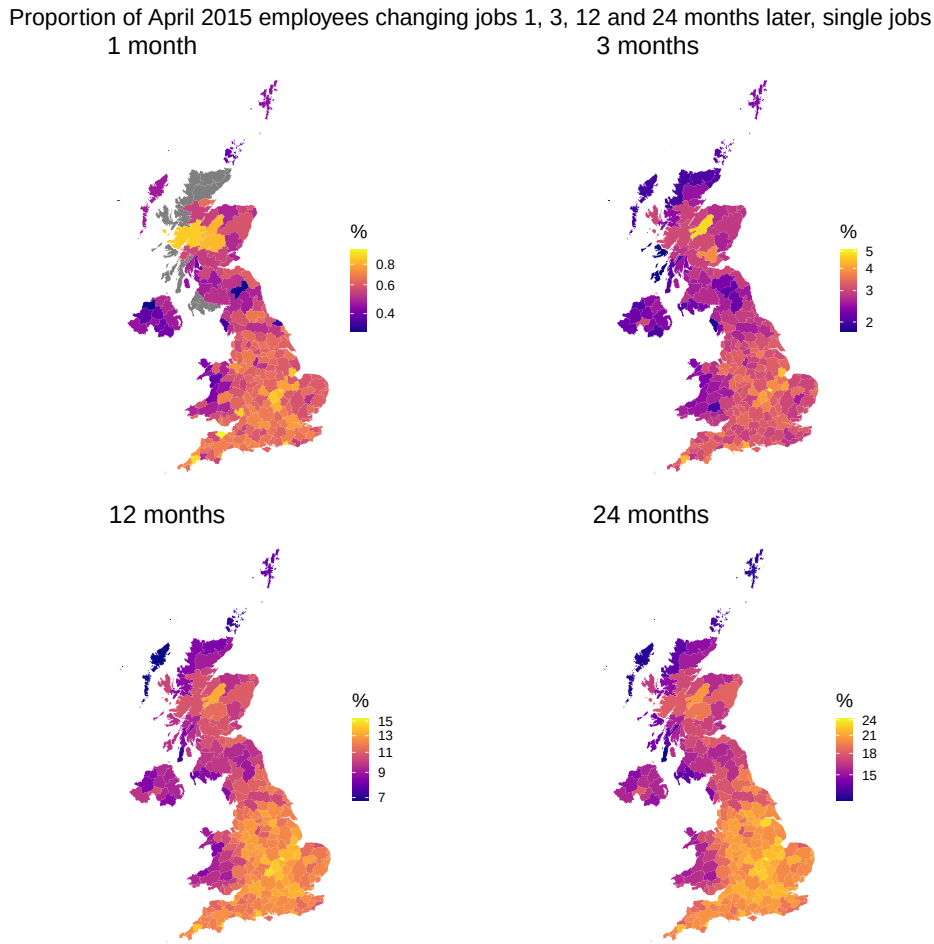


Figure 22
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in April 2015.

Figure 23 shows the case for individuals who change at least one of their multiple jobs while keeping the same number of jobs. One way this is illustrative is that there are numerous TTWAs where statistics cannot be produced due to the small number of observations falling below the disclosure threshold. We begin to reach the limits of what is possible, even using data on the whole population; this category is too small to permit comprehensive analysis of TTWA variation. Substantively, Figure 23 provides less evidence of a periphery effect than when considering single job holders.

Figure 24, which shows the proportion of employees taking on more jobs, suggests a quite mixed geographic pattern. Central and north west Scotland have a higher proportion taking on additional jobs. When considering 24-month results, there is perhaps also a higher tendency to

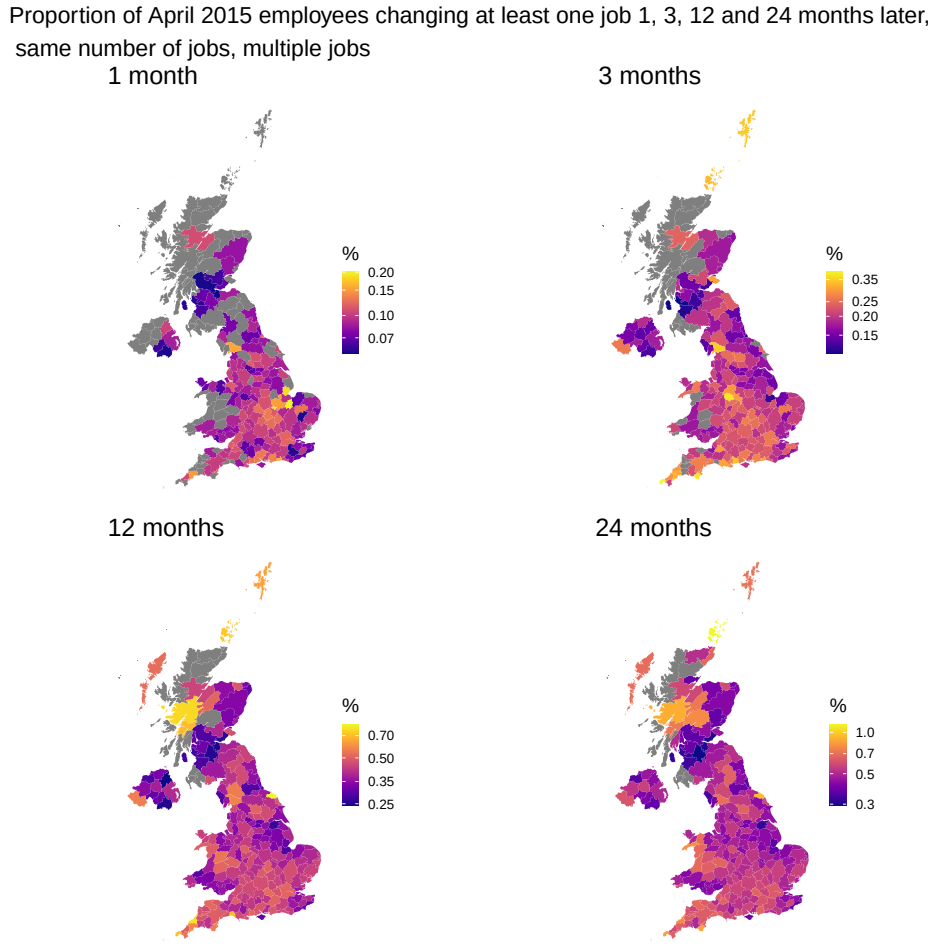


Figure 23
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in April 2015.

take on more jobs in the south west of England.

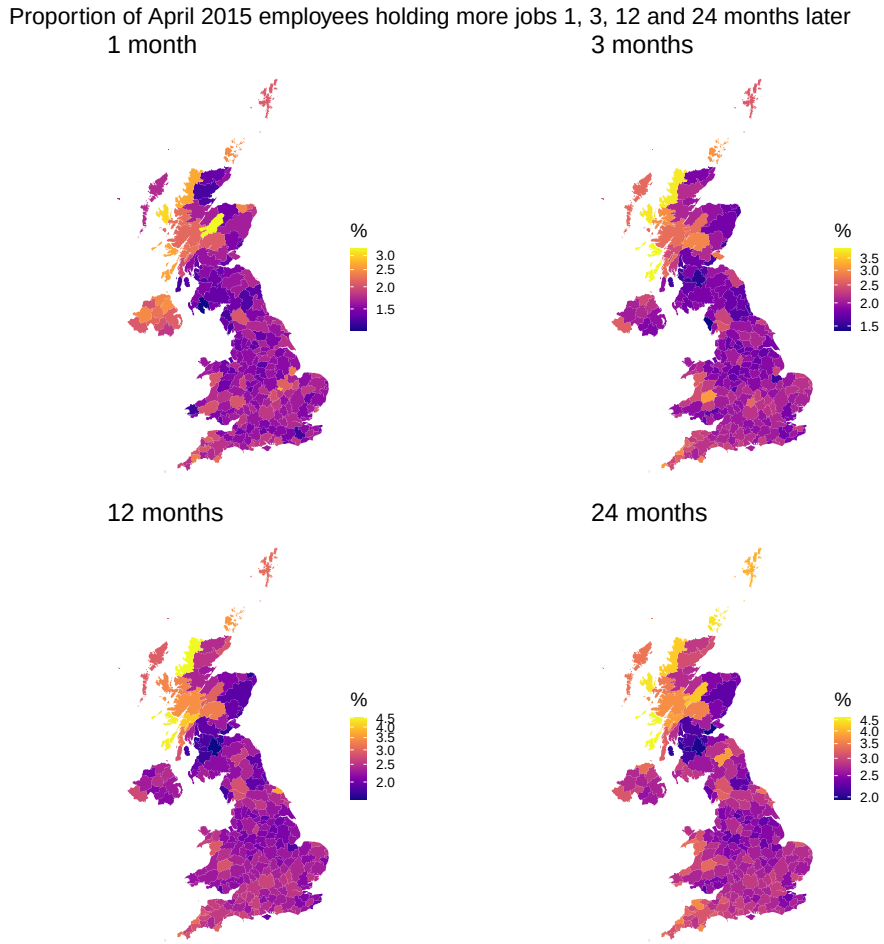


Figure 24
Source: authors' computation based on the 2014-2017 waves of the UK HMRC PAYE RTI data.
Notes: the breakdowns by geographies take the situation of the worker in April 2015.

5 Earnings transitions

This section presents statistics showing earnings mobility, as captured by changes over time in individuals' average payrolled weekly earnings. The earnings categories considered again highlight the granular analysis possible when using population data. We focus on changes over the 12 months from April 2015, as captured by movement between different categories. The first two categories are defined relative to the (age 25+) minimum wage; the first is the weekly pay equivalent of a part-time (17.5 hours) minimum wage job, the second is the weekly pay equivalent of a full-time (35 hours) minimum wage job. The third pay category ranges from above the full-time minimum wage to the 4th decile of the earnings distribution. Subsequent categories are the deciles of the earnings distribution of the whole population in April each year, up until the top decile which is is further divided in percentiles. Hence, it becomes possible to identify transitions among the top 1%, for example. Table 1 shows the earnings ranges associated with

each category for 2015 and 2016. Note that these categories are consistent across sub-groups.

April pay categories Weekly, £	2015	2016
100th percentile	> 2523	> 2580
99th percentile	1845 - 2523	1888 - 2580
98th percentile	1537 - 1843	1573 - 1886
97th percentile	1352 - 1536	1382 - 1572
96th percentile	1224 - 1351	1248 - 1381
95th percentile	1129 - 1223	1152 - 1247
94th percentile	1055 - 1128	1075 - 1151
93th percentile	994 - 1054	1015 - 1074
92th percentile	946 - 993	963 - 1014
91th percentile	906 - 945	923 - 962
9th decile	672 - 905	685 - 922
8th decile	541 - 671	552 - 684
7th decile	447 - 540	457 - 551
6th decile	371 - 446	381 - 456
5th decile	308 - 370	317 - 380
Part of 3d, 4th decile	235 - 307	235 - 316
< FT min wage	117 - 234	117 - 234
< PT min wage	< 117	< 117

Table 1: Weekly pay categories in 2015 and 2016

Notes: in 2015, the "< PT min wage" category is approximately equivalent to the first decile of the weekly pay distribution (<£119). The "< FT min wage" category comprises the second decile (£120-179) and most of the third (£180-243).

Figure 25 summarises transitions between the April 2015 (y-axis) and April 2016 (x-axis) earnings categories focusing on the deciles of the distribution. In addition, a non-payrolled category is included in April 2016. Including this category allows an insight into how the probability of being out of payrolled work varies with earnings. We note that this category will also include individuals who are not payrolled employees but are self-employed.

With regard to leaving payrolled employment, Figure 25 confirms the impression from Figure 2 that low earners in April 2015 are more likely than higher earners to be out of payrolled employment 12 months later; 24% of employees in the lowest earnings category are no longer in payrolled employment 12 months later. This compares with 13% for those in the second earnings category and successively smaller proportions up to the 9th decile. The higher proportion among the top decile is perhaps partly driven by the ability to retire early as the gap with other deciles increases after age 56 (see Figure 27 below).

Turning to earnings mobility, the initial overall impression is one of stability. In all 2015

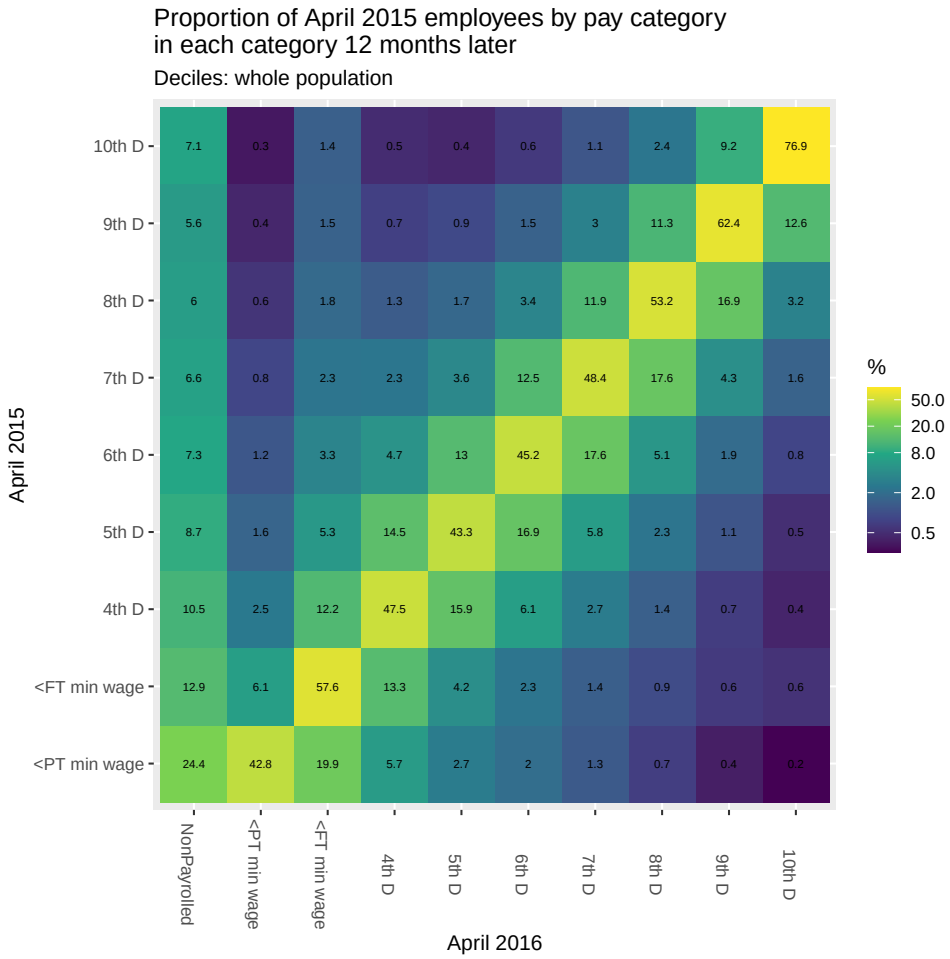


Figure 25

Source: authors' computation based on the 2015 and 2016 waves of the UK HMRC PAYE RTI data.
Notes: the weekly pay distributions are computed at the UK level in April 2015 and 2016. Upward mobility lies below the diagonal where lighter colours mean more mobility. All rows add to a hundred.

categories, most common is to remain in the same category in 2016 and less than 30% will move beyond a neighbouring category. Stability is lowest for employees below half time the minimum wage (43 %) due to transitions towards being non-payrolled and also 20% towards the next earnings category. Upward mobility is also comparatively high in the middle of the earnings' distribution, from the 5th to the 8th deciles. Stability is higher in the category below the minimum wage (58%) and highest in the 2 top deciles (62% and 77% respectively).

Figure 26 further splits the top decile's transitions in percentiles, at which point the probability of no longer being in payrolled employment increases quite steadily in each subsequent percentile from 6% to 11 %. This might be due to higher proportions fully turning to self-employment or changing life style where traditional work is not the main activity. Earning stability also increases with percentiles and is as high as 64 % for the top 1%. Note that this is stability within percentiles rather than deciles; a remarkable insight into mobility among top earners.

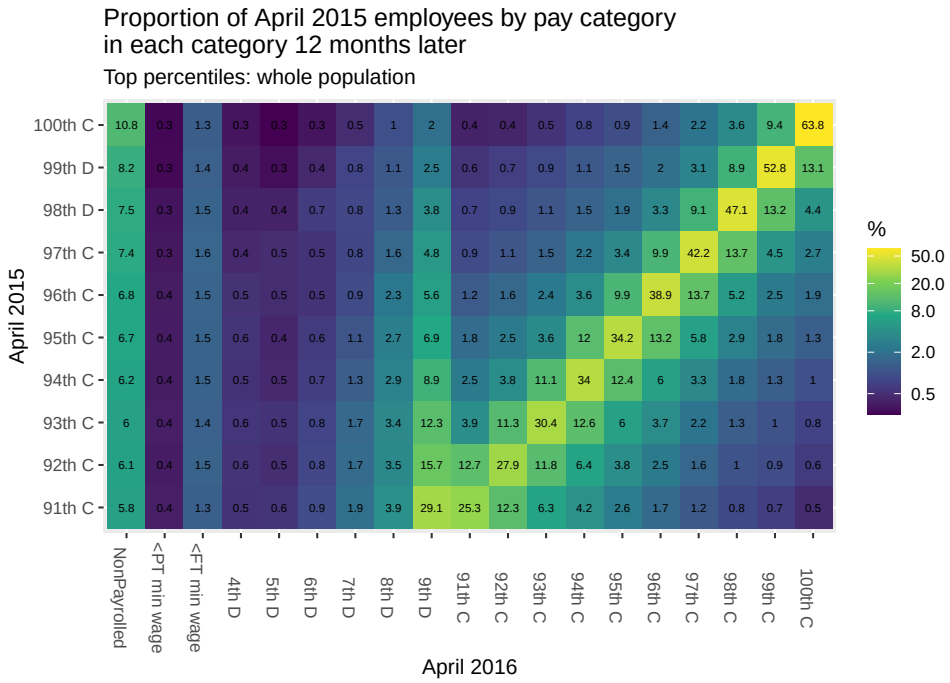


Figure 26
Source: authors' computation based on the 2015 and 2016 waves of the UK HMRC PAYE RTI data.
Notes: the weekly pay distributions are computed at the UK level in April 2015 and 2016. Upward mobility lies below the diagonal where lighter colours mean more mobility. All rows add to a hundred.

Figure 27 presents the same statistics for women and men, where we aggregate downward and upward mobility. While the overall impression is of similarity, there are some differences. For instance women consistently face greater earning stability, except for the top decile. This difference is particularly marked in the lowest categories. This indicates a greater tendency for women to remain in low earning payrolled employment. On the other hand, among men in the lowest earnings categories in April 2015, the proportion out of payrolled work 12 months later is larger than for women, e.g. 28% compared to 23% for the lowest category. Upward mobility is higher for men than women for all earnings in 2015, whereas the reverse is true for downward

mobility. A lack of progression of hourly wage or a decrease in the relative hours of work between men and women can explain this and we cannot distinguish between the two as we do not have the number of hours worked. The pattern within the top decile is more mixed. The top 1% of male earners are more likely to remain in the top 1% one year later than is the case for women (66% compared to 56%).

The final set of earnings transitions results show how mobility varies with age (Figure 28). We focus on employees aged 26-60 as the youngest and oldest age groups' status is influenced by education and retirement decisions, respectively. Differences are more pronounced here than across gender. For age groups 26-55, mobility among those below the top decile in 2015 is highest at younger ages and tends to reduce with age. The same applies to upward mobility and transitions out of payrolled employment. Interestingly, this is consistent across initial earnings level. Downward mobility is either unrelated with age or negatively related to it. Downward mobility and transitions out of payrolled employment increase in age group 56-60 compared to 46-55, probably linked to a decrease in hours worked and early retirement. This stability age profile is broadly maintained within the top decile. Being in the top percentile is particularly stable. One interpretation of this might be that there is a long tail of the earnings distribution that makes up the top 1% so individuals within this group can tolerate a substantial loss of earnings without it dislodging them into a lower category.

5.1 Sankey visualisations

Sankey graphs are an alternative visualisation which show flows out of and into payrolled employment together.

Figure 29 shows, for individuals employed in April, monthly flows into and out of payrolled work over the next year. The leftmost column corresponds to the population in payrolled employment in April 2015. Successive columns correspond to successive months of the tax year 2015. In each month, blue represents payrolled employment and green represents being out of payrolled employment. Hence, in April 2015 the column is entirely blue whereas in May 2015, by which time some people will no longer be in payrolled work, a proportion of the column is green. By June, more individuals have left payrolled employment but also some not employed in May have returned to payrolled work. To highlight such short-term transitions, the chart distinguishes employees who have been employed for only one month (lighter blue) from those who have been employed for longer (darker blue). Similarly, light green is used for individuals who are one month out of payrolled employment. We thus identify transitions over three months.

Since the charts take as their starting point the April population of payrolled employees, the numbers out of payrolled employment increase over time before stabilising in later months. It is frequent for employees to remain only one month in out of payrolled work; it encompasses 50% or more of the flows back towards payrolled employment for most of the months. On the contrary, periods of payrolled employment rarely last only one month.

Figure 30 shows the year-on-year mobility between 2014/15 and 2017/2018 of the average weekly pay over each year. Each column shows pay deciles for a single tax year: the leftmost (labelled '2014') relates to 2014/15; the rightmost, to 2017/18. Non-payrolled is defined as zero payrolled earnings over the full year. This does not exist as a category in 2014/15 since Figure

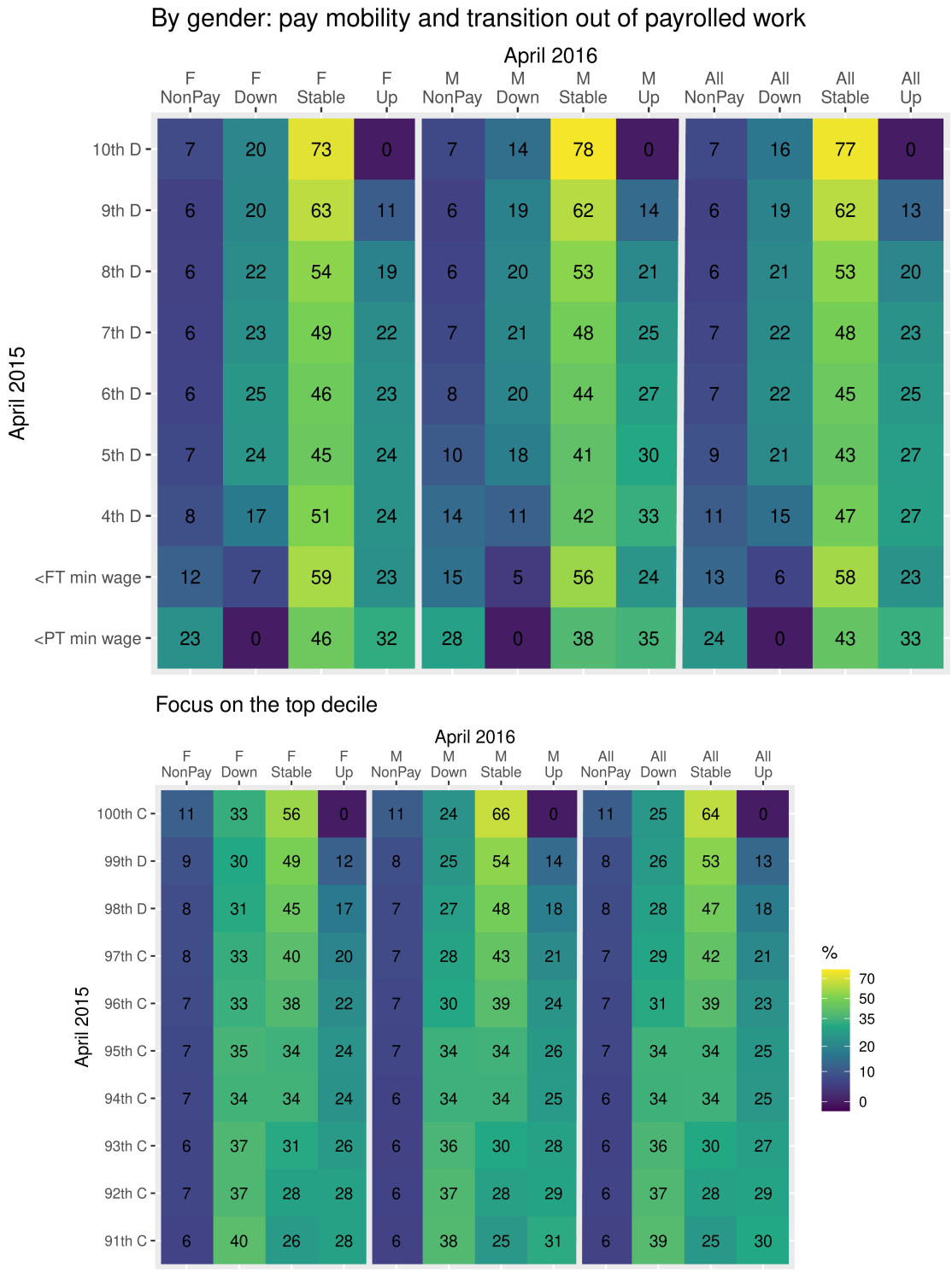


Figure 27: Proportion of April 2015 employees by pay category in each category 12 months later: by gender

Source: authors' computation based on the 2015 and 2016 waves of the UK HMRC PAYE RTI data.

Notes: the weekly pay distributions are computed at the UK level in April 2015 and 2016. The 4 columns show transitions out of payrolled work, downward mobility, stability and upward mobility. All rows by category add to a hundred.

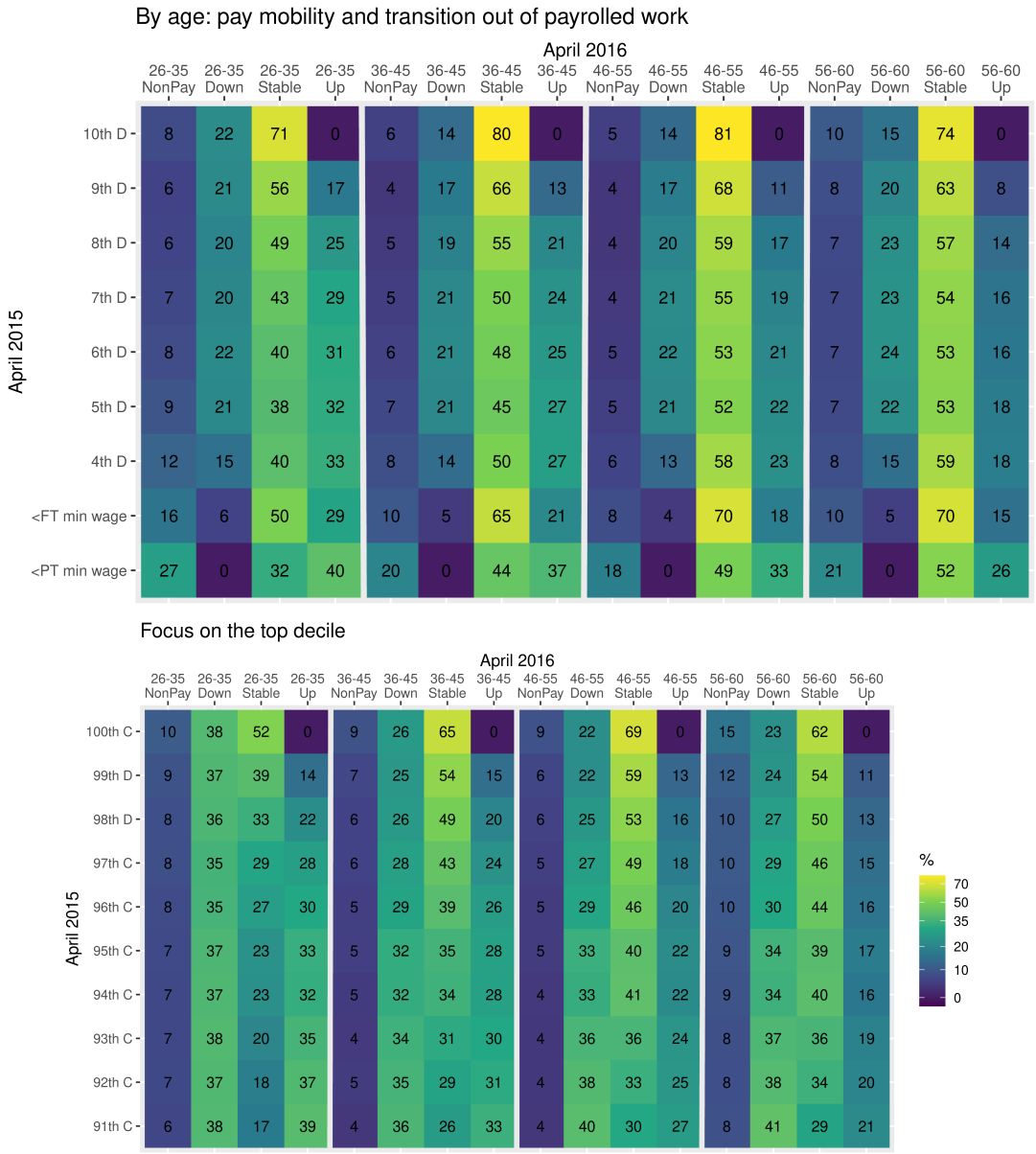


Figure 28: Proportion of April 2015 employees by pay category in each category 12 months later: by age

Source: authors' computation based on the 2015 and 2016 waves of the UK HMRC PAYE RTI data.

Notes: the weekly pay distributions are computed at the UK level in April 2015 and 2016. The 4 columns show transitions out of payrolled work, downward mobility, stability and upward mobility. All rows by category add to a hundred.

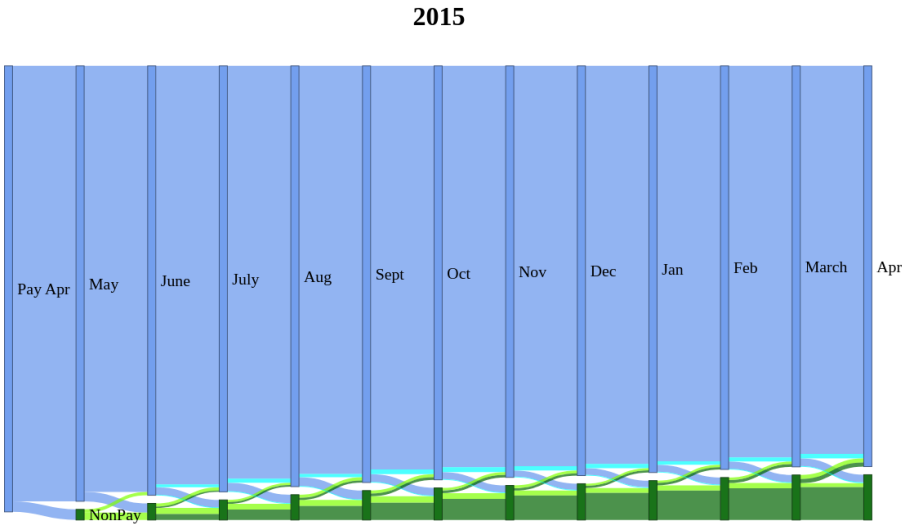


Figure 29: Payrolled employment flows
Source: authors' computation based on UK HMRC PAYE RTI data.
Notes: blue indicates payrolled employment, green indicates being out of payrolled employment. Light blue shows flows for payrolled employees who were out of payrolled work throughout the previous month. Light green shows flows for non-payrolled individuals who were payrolled during the previous month. Darker blue and green respectively show flows for those who have been payrolled or non-payrolled for more than one month.

30 is based on all individuals observed to be payrolled at some point in that year. However, of these, a proportion will have no earnings throughout one or more subsequent years. These are shown as non-payrolled in Figure 30. Because low paid employees are more likely to be non-payrolled some months in the year, there is a larger proportion of overall employees that are below the full-time equivalent of the hourly minimum wage when considering yearly income rather than monthly income, conditional on payroll employment. That is why the 4th decile is now absorbed within the minimum wage full-time equivalent category.

As an alternative perspective to 29, the flows in 30 are coloured to highlight the destination state rather than the origin state. We see that low paid employees are highly likely to also have been low paid in the previous year. Likewise, those out of payrolled employment are predominantly drawn from those who were out of paid work or on low pay in the previous year. Yearly pay mobility is highest in the middle of the distribution. For instance, between 2014 and 2015, 43% of employees in the 6th decile remained there in the tax year 2015, whereas 66% remained in the 9th decile, 81% in the top decile and 61% below the full-time minimum wage.

6 Conclusion

This paper has provided some novel statistics intended to demonstrate the analytical potential offered by HMRC PAYE RTI data. While these data are already being used to produce some labour market statistics, numerous opportunities exist to develop their use further. In this paper we have focused on the longitudinal nature of the data and presented results on individuals'

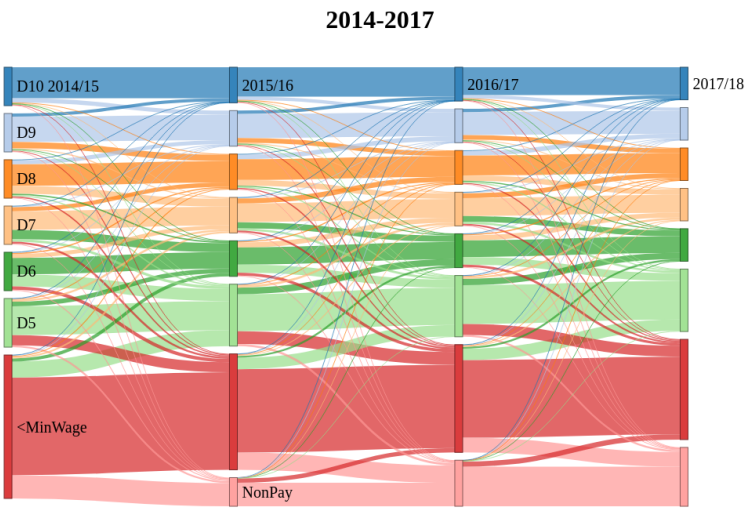


Figure 30: Annual pay mobility - movement between earnings deciles
Source: authors' computation based on the 2014-2018 waves of the UK HMRC PAYE RTI data.
Notes: weekly pay average for each year. Non-payrolled means that pay was zero for the whole year. The 4 columns are each tax year. D5 to D10 are decile 5 to decile 10 of the pay distribution, computed every year, where the 5th decile category also comprises part of the 4d decile above the full-time equivalent of the hourly minimum wage. The last pay category refers to pay below that number. The hourly rate for the National Minimum Wage refer to the one for 21 years old + in April each year.

payrolled employment, job and earnings transitions. We have shown how these vary according to individuals' characteristics, locations and circumstances. Of course, simple descriptive statistics do not capture the extent to which variation on one dimension is likely to interact with variation on another dimension. In principle, the PAYE RTI data could straightforwardly be used to conduct a multivariate analysis that would allow for such interaction.

The paper has highlighted several key advantages of the data. Most obviously, their large size means robust statistics become possible for small population subgroups. This has been illustrated through the production of statistics showing TTWA-level variation in payrolled employment transitions among population subgroups, for example. To the best of our knowledge, such statistics have not been produced before. It is simply thanks to the size of the data that these detailed insights into the labour market become possible.

A second advantage that is perhaps less immediately obvious is that, since the data cover the population of payrolled employees, they can also be used to construct variables characterising the population of employing firms. We made use of this in the paper by constructing a variable of firm size in terms of number of employees. Much more could be done here; since jobs are associated with employers, the data provide the basis for a linked employer-employee dataset. Further linkage with, for example, the IDBR could develop the data in this direction. Similarly, person-level data could in principle be linked with other individual-level administrative data to enrich worker information. Matching in small area information is also possible. For instance, scraped online vacancy data at the TTWA would allow the extent of local labour mismatch to be directly assessed.

The data have other significant strengths. They are available at high-frequency and, since

they are not reliant on the cooperation of subjects, are not complicated by the issues that beset surveys: non-response, recall bias, high cost. Furthermore, they permit individuals to be observed over a long period. This paper used data across four tax years. Over time, the longitudinal dimension will naturally extend and will provide a unique resource for the observation and analysis of long-term trends. It is worth noting that this will tend to reduce the limitation of the data being restricted to the employed population only. Over time, the subgroup of the population that is not observed will reduce to those who have never been in payrolled. Hence, rather than capturing the employee population, the data will increasingly come to resemble the population of the economically active. This will not be fully achieved – for instance, individuals who worked at some point and subsequently left the labour market cannot be distinguished, without additional information, from former workers who are searching for a job – but nevertheless it may provide a stronger basis for examining transitions in and out of payrolled employment.

As a final comment, it should be emphasised that there remains a significant role for survey data in the production of labour market statistics. While survey data have their own difficulties, they offer advantages over administrative data in several regards. In some cases, the limitations of the administrative data could potentially be addressed. For instance, unlike the LFS, the PAYE RTI data do not cover self-employed people or (explicitly) those out of work; linking to self-assessment tax data and benefit records could improve this. However, other limitations are more fundamental. In particular, surveys can collect information on the precise outcomes of interest, rather than be limited to what is collected for administrative purposes. The PAYE RTI data are lacking in this regard. For example, they lack detailed information on hours of work, occupation or the background characteristics of workers. These relative strengths of survey data are sufficient to ensure the continued relevance of survey-based statistics. Rather than being a substitute for survey data as the basis for producing labour market statistics, administrative data provide a complement.

Appendix A: Constructing the dataset

1 Introduction

This note sets out the processing approach that has been applied to HMRC PAYE RTI data in order to construct a dataset suitable for the production of statistics on labour market transitions. These transitions may be between jobs, between employment and non-employment or between earnings levels.

The new statistics are intended to provide fresh insights into the nature and functioning of the labour market. To do this, requires a re-purposing of the PAYE RTI data. The supplied data record monthly earnings notified to HMRC. However, it is not always appropriate to interpret recorded earnings in a month as corresponding to employment in that month, nor the absence of recorded earnings as non-employment. A disparity would arise, for instance, where an employer was late in submitting a return to HMRC.

Such recording issues are relevant for statistics on labour market transitions since they can create a misleading impression of job, employment or earnings mobility. A key aim of the data processing is to impute employment from earnings in a sensible way, smoothing spells when appropriate but not losing true breaks in jobs. The processing rules have been developed after inspecting numerous cases within the data. Despite this, it is not possible with the available data to know how well they have achieved their aims and the resulting statistics should be caveated to this effect. Econometric models may be able to incorporate such uncertainty more directly.

The next section of this document describes the data. The aim of the data processing is described in section 3 and the processing carried out to achieve this is presented in section 4. The resulting dataset is assessed in section 5 by comparing it with other published data. Section 6 sets out the remaining programme of research using the new data.

2 Characteristics of the data supplied

The data for this project cover the tax years 2014/15, 2015/16, 2016/17 and 2017/18. Four separate earnings files were supplied, one for each year. Records relating to pensions were removed from the data following the approached described in the HMRC experimental statistics (using the OccPen_signal variable).[1] Each of the remaining records relates to a job.

The variables used in the analysis were:

- **Person identifier**: *nino_anon*
- **Job identifier**: *emp_no* (in conjunction with *nino_anon*, this identifies each job for an individual).
- **Job start and end dates**:
  - *jobstart*
  - *jobend*
- **Person characteristics**: sex

[1] HMRC (2019) Earnings and Employment Statistics from Pay As You Earn Real Time Information: Experimental Statistics, April 2014 to September 2018.

- **Earnings variables**:
    - *mpay_YYYYMM* – earnings received in a given month
    - *payment_frequency_YYYY* – the period each payment covers
    - *N_payments_YYYYMM* – the number of payments received in a given month.

Additional files were supplied to augment the earnings data with the following variables:

- **Employer identifier**:
    - *entref* – enterprise identifier
    - *wowref* – who owes whom identifier
    - SIC 2007
- **Person characteristics**:
    - month of birth
    - TTWA
    - LA
    - region

The key variable for our analysis is monthly earnings and the definition adheres to that used in the HMRC experimental statistics. It is constructed as taxable pay plus contributions to occupational pension schemes, and the element of payroll giving (charitable donations) and the value of childcare and other non-cash vouchers which are assessed for National Insurance Contributions. Benefits in kind are excluded. This definition is intended to represent as closely as possible the measure of headline gross pay that employees see on their pay slips (and so is consistent with a common understanding of total pay).

3 Aim of the data processing

Data processing is carried out in order to construct a 48-monthly panel dataset. There are several challenges involved in preparing the data for analysis. Conceptually, there is the question of how we infer employment and wages from the PAYE RTI data. Practically, there is the challenge of imposing this conceptualisation on the data. In this section, both aspects are discussed. Details of the business rules adopted are the focus of the next section.

3.1 Employment

We infer employment from positive earnings. Start- and end-dates of jobs are recorded. Ideally, these would indicate the months during which the job existed. In practice, however, observed earnings are sometimes inconsistent with the dates provided (an indication of the extent of this is provided in the next section). Payments may be observed before the start of the job and/or after the end of the job. An alternative to relying on supplied dates is to simply regard months with positive earnings as months of employment. The difficulty with doing that is that there will sometimes be months with no earnings, despite the start- and end-dates indicating the job was live at that time. This may be for observable reasons – perhaps the job is paid less frequently than monthly – but may also be for reasons that are unobserved such as the employer return to HMRC being late, the amount compensating for an earlier error, or the job being on a zero-hours contract or other less regular basis.

The approach to imputation described below sought to identify plausible business rules that could be applied to the data in order to capture the nature of the underlying job spell. If a break in earnings appears to arise in the course of a job spell for reasons that are unlikely to indicate that the job truly stopped for a period, it is appropriate to impute the job spell as being unbroken. To do otherwise risks over-estimating the number of transitions into and out of jobs, which is particularly undesirable given the focus of this project.

This imputation approach was informed by inspecting numerous individual job spells. In doing this, natural imputations would often suggest themselves. However, the number of records involved is such that tailoring the approach to individual cases is not feasible. Rather, the challenge was to identify business rules that were generally applicable, recognising that they may not be appropriate in all cases.

3.2 Earnings

Inferring employment from positive earnings in a month is conceptually straightforward. However, there is variation across jobs in the period to which earnings within a month relate. This feature complicates the interpretation of monthly earnings. Our aim is to process the data in order to render earnings amounts more directly comparable across jobs. For instance, similarly-paid jobs will have an earnings profile that is more lumpy when paid quarterly than when paid monthly. Analogously, jobs paid weekly will appear less well paid if they have only existed for part of the month. Ideally, an hourly rate of pay would be available to allow jobs to be directly compared. Unfortunately, the hours variable included in the PAYE RTI data is banded and of low quality and so does not allow this. Instead, the approach adopted is to convert earnings within a month into an implied weekly rate of pay. This uses information on payment frequency and on the number of payments received within a month. This method is imperfect to the extent that it does not deal with jobs involving one-off or irregular payments (about 1% of all jobs) nor does it reflect reduced earnings arising from only part of the payment period being worked. However, it does improve comparability of compensation level across jobs and is therefore better suited to the aim of comparing earnings progression, for example.

The level of earnings informs the choice of approach to dealing with breaks in employment. In some cases, the appropriate step to deal with such a break will be reasonably clear on seeing the earnings levels. For instance, a one-month break in recorded monthly earnings followed by a payment of twice the usual size suggests that half of the double payment should be moved to the month showing no earnings.

The earnings data present the additional complication that some months are recorded as having negative earnings. Again, the earnings level is informative. In many of the cases examined, negative payments offset earlier positive payments, suggesting the negative payment to be a correction. In such cases, the appropriate processing rule follows naturally.

4 Data processing

4.1 Processing in R (mkdata.r)

The data were supplied as R .fst files. These were read into R and saved as Stata format .dta files. Data processing was carried out in Stata since this proved faster than R for some data processing operations. Processing in R was confined to the applying the following initial steps to the supplied .fst files:

1. Drop entries that relate to pensions
2. Save in Stata format

4.2 Processing in Stata (processdata.do)

4.2.1 Initial processing

First, the Stata files saved in R were read into Stata. These files had more precision than necessary (pay amounts may have more than two decimal places, despite only 2 being needed). We rounded to two decimal places which then allows data to be held more efficiently.

Next, we identified job records that were duplicates in respect of job identifiers and monthly earnings variables. Duplicates were dropped (that is, a single instance of the record was kept).

The annual datasets were then merged to create a single 48-month job-level dataset. Jobs with zero total earnings over the 48 months were removed. The resulting file has records for 86,834,815 jobs.

4.2.2 The need for imputation

Working with data on this scale is slow, so imputation rules were developed using a sample of roughly half a million jobs. In constructing monthly series, we focus particularly on those jobs paid on a weekly, 2-weekly, 4-weekly or monthly basis. These account for nearly 99% of recorded jobs in all years, as evident from Table 1.

Table 1 Payment frequency by tax year

	2015	2016	2017	2018
One-off	0.1%	0.1%	0.1%	0.1%
Irregular	0.9%	1.0%	0.9%	0.8%
Monthly	65.3%	65.6%	65.9%	66.1%
3-monthly	0.0%	0.0%	0.0%	0.0%
6-monthly	0.0%	0.0%	0.0%	0.0%
Annually	0.1%	0.1%	0.1%	0.1%
Weekly	23.1%	22.9%	22.8%	22.6%
2-weekly	2.7%	2.7%	2.7%	2.7%
4-weekly	7.7%	7.6%	7.5%	7.4%
Number of records	251277	265276	271340	276997

The degree of inconsistent earnings records is illustrated in Table 2, in which consideration is further narrowed to those jobs paid at least monthly (as evident from Table 1, this results in negligible sample loss).

Table 2 Inconsistencies in the spells data

Earnings observed before recorded job start	0.8%
Job recorded as starting within 2015-18 but first earnings not observed until 2+ months later	3.0%
Job start variable suggests job start pre 2015 but first earnings not observed until later	3.0%
Job recorded as ending within 2015-18 but last earnings observed 2+ months earlier	4.7%
Job recorded as ending within 2015-18 but last earnings observed 2+ months later	2.6%
Job should continue beyond 2018 but last earnings observed earlier	8.3%

Base = 519,673

4.2.3 Business rules

Negative earnings

Business Rule 1: if month t has negative earnings, add these to t-1 and set t earnings to 0. If, after doing this, month t-1 has negative earnings, repeat the process using month t-2 and continue in this way until the negative amount disappears or as far back as possible. After processing, drop jobs with no positive earnings and set to zero negative earnings preceding positive earnings.

Negative earnings can arise as a correction where an individual has previously been overpaid. Negative payment amounts are often seen in a single month following a sequence of positive monthly earnings amounts. In this case, the interpretation as a correction is natural, and the remedy is to add the negative amount to the last positive amount. However, negative earnings amounts can also arise in ways that are less straightforward to rationalise. For instance, they can arise before an employment spell has started. Or they can arise during employment spells.

Of the 519,673 jobs in the sample file, 2.1% have a negative entry in at least one month. Inspecting a number of cases revealed that it was common for the negative amount to be equal in size to the preceding (positive) amount. In some cases, the negative amount was equal in size to the sum of a number of earlier positive amounts.

In view of this, the approach adopted to deal with negative entries was to add them to the positive amounts in previous months for as many months as require to fully absorb the negative amount. In a small number of cases, this ended up with the job showing no positive amount. This might be due to cancelling out or to the negative amount being (absolutely) greater than the sum of all earlier amounts.

Figure 1 provides some illustration:

- With Job 1, the original data (left panel) shows a negative amount in month 3 of equal size to the (positive) amount in month 2. Interpreting this as a correction, the imputation procedure results in months 2 and 3 cancelling out (right panel).
- With Job 2, the negative amount in month 4 is greater than the amount in month 3. After adding month 4 to month 3, the new month 3 is negative and so in turn is added to month 2. The net effect is to cancel months 2-4.
- With Job 3, the month 4 amount is of equal size to the preceding 3 months so the imputation rule results in cancelling all entries.
- With Job 4, the month 4 amount cannot be reabsorbed and applying the imputation rule leaves a negative amount in the imputed series in month 1.
- With Job 5, the negative amount in month 4 is smaller (in absolute terms) than the amount in month 3, so month 3 earnings are reduced by the month 4 amount and month 4 earnings are set to zero.
- With Job 6, the negative amount in month 4 is larger (in absolute terms) than the amount in month 3 but smaller than the total amounts in months 2 and 3. Adding the negative month 4 amount to month 3 leaves month 3 earnings negative. This new month 3 negative amount is then added to month 2. The net result is that month 2 earnings is reduced by the absolute difference between the month 3 and month 4 amounts, and earnings in months 3 and 4 are then set to zero.
- Lastly, with Job 7, the negative amount in month 4 can be absorbed by the month 3 amount while still leaving a positive amount in months 1-3 in the imputed series.

Figure 1 Absorbing negative amounts

	Original				**imputed**			
month:	1	2	3	4	1	2	3	4
job 1	x	x	-x	x	x			x
job 2	x	x	x	-2x	x			
job 3	x	x	x	-3x				
job 4	x	x	x	-4x	-x			
job 5	x	x	x	-y	x	x	x-y	
job 6	x	x	x	-z	x	x-(z-x)		
job 7	x	x	2x	-x	x	x	x	

In practice, most negative amounts can be absorbed while still leaving positive earnings amounts in at least one other month. Of the 519,673 jobs in the test sample, only 0.1% had no positive earnings post-imputation. These cases were then dropped, reducing the test sample to 519,171. Jobs whose first non-zero amounts were negative but were later followed by positive amounts had their negative entries set to zero. They were few in number (0.01%) and the choice to retain them was based on the rationale that the negative amounts all preceded positive amounts and so may be corrections for over-payment in an earlier period.

Constructing a measure of weekly earnings

Business Rule 2: calculate weekly earnings by dividing monthly earnings by the number of weeks each payment covers (as implied by the recorded payment frequency multiplied by the number of payments in that month). Note: this is not done for those receiving one-off or irregular payments.

We use the measure of earnings within a month, payment frequency (as presented in Table 1) and the number of payments received within a month to derive a measure of weekly pay. Doing this is helpful to the extent that it allows compensation comparisons across jobs and workers. It also permits comparisons with the Average Weekly Earnings official statistics.

It is important to be aware of how the measure is constructed in order to interpret it appropriately:

- No weekly measure is constructed for jobs where the payment frequency is "one-off" or "irregular" (nor where it is missing)
- There is no guarantee that the individual worked throughout the payment period. For instance, someone paid monthly may have worked for only some of the weeks in that month. This may arise most obviously in the case of starting or ending a job part way through the month. The imputation approach does not address this.

Creating an employer identifier

Before applying further business rules, we first create an employer identifier. The *nino emp_no* combination identifies separate jobs for an individual but does not constitute an employer identifier. One reason for wanting this is in order to ignore transitions between enterprises within the same corporate group. We create a new identifier – *employer* – that is the same as *entref* unless there is a linked *wowref* in which case it takes that value. Where individuals have multiple records with the same employer, we consolidate these into a single record.

Broken spells

Business Rule 3: Where there is an earnings gap of 1 month and the month before (after) the gap is double that after (before) the gap, earnings in the double month is shared equally with the missing month.

Business Rule 4: Where earnings before (after) a missing month is double that of the month before (after), earnings in the double month is shared equally with the missing month.

Business Rule 5: Where the earnings amount in one month is a higher integer multiple, M, of the following month, and there are M-1 preceding months with no earnings, that multiple amount is shared with these preceding months (unless this would affect months prior to the recorded start month of the job).

Business Rule 6: Where a sequence of at least two successive non-missing months is followed by missing months and then a one-off payment, this one-off payment is moved (added) to the last month of the continuous sequence.

The appropriate steps to deal with these broken spells depends on the likely cause. The challenge is to distinguish possible reporting issues from true transitions so that a reasonable imputation rule can be used to impute employment.

In some cases, the appropriate rule is relatively clear:

- in the case of a **zero-hours contract**, the individual may have an ongoing contract of employment but will only receive payments during the months they are productive. In this case, it is not appropriate to view the individual as employed during non-productive months.
- in the case of **errors in payment**, the individual may be productive despite payment being delayed. Inspection of the data reveals cases where the earnings amounts are suggestive of a correction for earlier months where there are no earnings. In such cases, it is plausible to view the months prompting the correction to be productive and the individual to be employed.

The true reason for a break in payment is not known. The business rules developed to impute breaks are discussed below. The aim is not to remove all gaps since they may represent true breaks in employment. Instead, the rules intend to strike a balance between bridging gaps that are likely to be the result of mis-recording vs smoothing over true breaks. After applying these rules, the proportion of jobs showing a break was 14.5%.

Business Rule 3: There are instances where the sequence of earnings amounts has a missing month but the amount before the missing month is twice that after the break (or vice versa). In such cases, the approach is to share the earnings amount in the "double" month equally with the missing month. This is illustrated in Figure 2. The first row corresponds to the case of the double month preceding the missing month, while the second row corresponds to the case of the double month coming after the missing month.[2]

Business Rule 4: There are instances where the payment before a missing month is double that of the month before. Alternatively, the payment after a missing month may be twice that of the subsequent payment. These scenarios are illustrated in Figure 2. They differ

[2] In practice, a tolerance is allowed – earnings in the double month must be within 1% of earnings in the regular month. A similar tolerance is used throughout. When filling longer gaps, the tolerance is multiplied by the number of months being imputed.

from Business Rule 3 in that it is no longer necessarily about bridging a gap (the missing month does not need to have neighbouring non-missing months). Again, the double amount is shared with the subsequent or preceding month as appropriate.

Business Rule 5: There are instances where the earnings amount in one month is a higher multiple of the following month. Where this multiple is matched by the number of preceding months with no earnings, that higher-multiple amount is spread across the matching number of months. This is shown in Figure 2 for the case of a gap (that is, there is a non-missing month preceding the missing months). This need not be the case. Note also that the rule is not applied to those cases where the imputation would affect months prior to the recorded start month of the job.

Business Rule 6: After this processing, there remain cases where a sequence of non-missing months is followed by missing months and then a one-off payment. Such "erratics" are often small amounts and seem likely to be balancing items. We deal with these by consolidating them with the unbroken spell, adding them to the last observed payment in the continuous monthly sequence. In implementing this rule, we require only that the continuous spell last at least two successive months.

Schematically, the approach is shown below. Note that Business Rule 6 is shown for two consecutive missing months but the imputation approach is general and can deal with any length of missing.

Figure 2 Imputing broken spells

	original				**imputed**			
	1	2	3	4	1	2	3	4
Rule 3	x	2y		y	x	y	y	y
	x		2x	y	x	x	x	y
Rule 4	x	2x			x	x	x	
		2y	y		y	y	y	
Rule 5	x			3y	x	y	y	y
Rule 6	x			y	x+y			

5 Assessing the results

The performance of the processing was assessed in two ways. First, a small dataset was produced that listed pre- and post-processing data for a sample of cases selected to cover the range of imputations used. This dataset was inspected on a case-by-case basis to check that the results looked sensible and the rules were being applied in the intended way. The result of this inspection was positive in this regard.

The second check on the data involved comparing statistics based on the processed data with analogous published statistics.[3] The results of doing this are presented in this section and involve the following comparisons:

- Employee numbers, compared against
  - HMRC experimental stats (UK_Real_Time_Information_Experimental_Statistical_Tables_1a_-_3b__-_April-2014_to_December-2018.xlsx, tab "table 1a")
  - LFS equivalents (emp01nsamay2019.xls, column c)
- Earnings, compared against
  - HMRC experimental stats (UK_Real_Time_Information_Experimental_Statistical_Tables_1a_-_3b__-_April-2014_to_December-2018.xlsx, tab "table 1b")
  - Average Weekly Earnings (AWE) equivalents (earn02may2019.xls, tab 4, column b)
- Gross flows into and out of employment, compared against
  - against LFS equivalents (x02may2019.xls, tab "Labour market flows NSA", columns e, f)
- Job-to-job flows, compared against
  - LFS equivalents (x02may2019.xls, tab "Labour market flows NSA", columns j, k).

5.1 Employee numbers

Error! Reference source not found. shows the estimated number of employees in the UK implied by the processed data, compared with published HMRC equivalents and with the LFS. The processed data suggest employee numbers very similar to those published by HMRC. The fact that the processed data record employee numbers consistently slightly below those presented in the published HMRC experimental statistics is due in part to the processed data including only those jobs for which weekly pay can be imputed. This excludes one-off jobs and jobs payed irregularly or less than monthly.

Figure 3 also allows a comparison with employee numbers as estimated from the LFS. The PAYE RTI data show considerably higher employment levels than the LFS. The HMRC experimental statistics report includes a detailed comparison of the data sources and a discussion of why such a difference exists. Partly, this is due to the greater coverage of the PAYE RTI data capturing more jobs. Partly, it is due to definitional differences between the data sources. A side-by-side comparison is reproduced below for convenience.

[3] The employment and earnings statistics are compared against the HMRC statistics published in 2019 (see footnote 1) since these were the most recent available at the time the analysis was conducted. The use of PAYE RTI data has evolved since then. For instance, statistics on flows were not produced in 2019 but are included in more recent releases (see https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/earningsandemploymentfrompayasyouearnrealtimeinformationuk/november2021).

Figure 3 Estimated employee numbers in the UK, 2013/14-2017/18 ('000s)

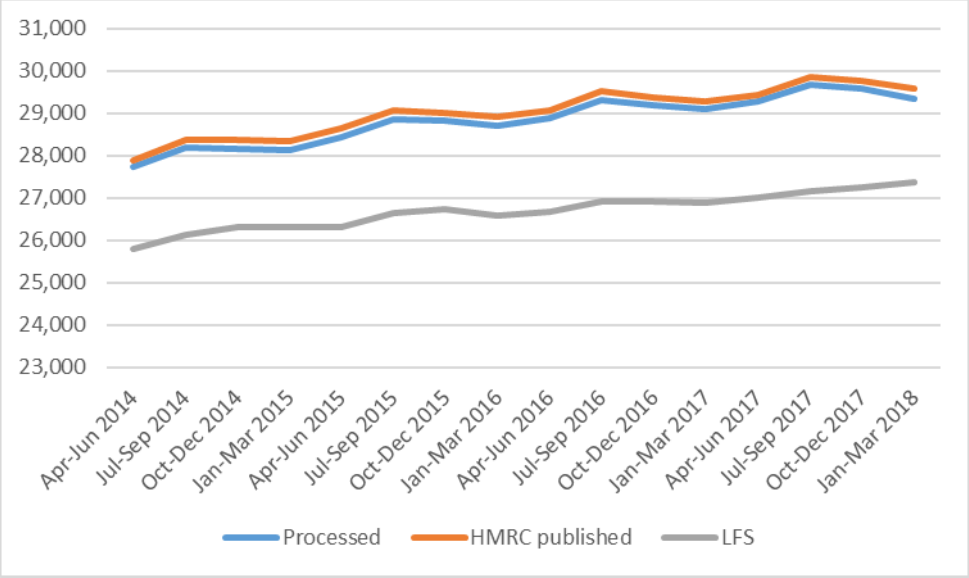


Figure 4 Side-by-side comparison of LFS and PAYE RTI data

	Labour Force Survey (LFS)	**Real Time Information (RTI)**
Timeliness	Published monthly. A six to seven week gap between the end of the reference period and the publication date.	Initial Publication Early 2018 as an experimental statistic. Subsequent publication schedule to be confirmed at a later date.
Employment measure	Anyone carrying out at least one hour's paid work in the reference week	Anyone who has received pay from PAYE in the reference period
Reference period	One week	One Quarter
Inclusions	UK resident population in • Private households • NHS accommodation • Young people living away from the parental home in a student hall of residence or similar institution during term time	All individuals being receiving pay through a PAYE scheme. This will include: • People living in communal establishments • Some foreign residents • People aged under 16
Exclusions	• Employees not paid during the reference period, e.g. for certain types of seasonal work (summer jobs or Christmas temps, for example) •Under 16s • Communal establishments such as residential care homes, prisons or defence establishments • Foreign residents	• Employed individuals in the undeclared economy whose income is not reported to HMRC via PAYE • Self-employed
Full-time / part-time breakdown	LFS includes data for all employees, full-time employees and part-time employees separately.	RTI does not differentiate between full-time and part-time workers.
Statistical adjustments	Non-response – imputation or rolling forward previous data from the respondent and reweighting of responses.	Late return grossing for open tax years.
Sampling variability	For employments aged 16+ ± 171 thousand for September to November 2017	Not Applicable

Table 3 Employee numbers (`000s)

Year:	2014			2015				2016				2017				2018
Quarter ending:	Jun	Sep	Dec	Mar	Jun	Sep	Dec	Mar	Jun	Sep	Dec	Mar	Jun	Sep	Dec	Mar
Processed data																
United Kingdom	27,730	28,197	28,161	28,130	28,437	28,852	28,821	28,714	28,892	29,315	29,194	29,088	29,276	29,670	29,591	29,337
England	23,279	23,675	23,670	23,661	23,906	24,254	24,246	24,168	24,296	24,646	24,548	24,475	24,614	24,943	24,887	24,662
North East	1,064	1,075	1,077	1,074	1,081	1,091	1,090	1,082	1,087	1,098	1,095	1,088	1,092	1,104	1,105	1,093
North West	2,995	3,043	3,048	3,040	3,066	3,106	3,109	3,092	3,106	3,145	3,142	3,125	3,147	3,190	3,192	3,160
Yorkshire and The Humber	2,209	2,253	2,255	2,244	2,272	2,301	2,301	2,289	2,302	2,330	2,324	2,311	2,323	2,353	2,352	2,328
East Midlands	2,023	2,055	2,055	2,051	2,071	2,101	2,103	2,091	2,100	2,132	2,128	2,116	2,127	2,156	2,155	2,130
West Midlands	2,362	2,402	2,404	2,403	2,429	2,466	2,472	2,456	2,473	2,507	2,506	2,492	2,515	2,549	2,550	2,520
East	2,637	2,678	2,673	2,673	2,697	2,735	2,731	2,723	2,732	2,777	2,761	2,754	2,766	2,802	2,792	2,768
London	3,757	3,838	3,862	3,885	3,935	4,010	4,034	4,044	4,066	4,132	4,128	4,140	4,156	4,217	4,221	4,197
South East	3,900	3,963	3,942	3,941	3,978	4,036	4,016	4,009	4,029	4,090	4,054	4,049	4,069	4,122	4,093	4,062
South West	2,332	2,369	2,354	2,350	2,377	2,408	2,390	2,383	2,402	2,435	2,409	2,399	2,418	2,450	2,426	2,405
Wales	1,221	1,238	1,234	1,230	1,241	1,257	1,253	1,246	1,255	1,271	1,265	1,258	1,268	1,284	1,280	1,265
Scotland	2,380	2,418	2,401	2,387	2,418	2,444	2,429	2,408	2,431	2,457	2,436	2,415	2,441	2,468	2,450	2,421
Northern Ireland	697	708	708	708	718	726	726	723	729	738	740	739	746	756	756	753
Published HMRC data																
United Kingdom	27,880	28,380	28,370	28,330	28,640	29,070	29,020	28,920	29,080	29,510	29,380	29,270	29,440	29,850	29,780	29,580
England	23,530	23,960	23,980	23,960	24,210	24,590	24,560	24,490	24,620	25,000	24,890	24,810	24,940	25,290	25,240	25,080
North East	1,070	1,080	1,090	1,080	1,090	1,100	1,100	1,090	1,100	1,110	1,100	1,100	1,100	1,110	1,110	1,100
North West	3,020	3,070	3,080	3,070	3,100	3,140	3,140	3,130	3,140	3,180	3,180	3,160	3,180	3,230	3,230	3,210
Yorkshire and The Humber	2,220	2,280	2,280	2,270	2,290	2,320	2,320	2,310	2,320	2,350	2,340	2,330	2,340	2,370	2,370	2,350
East Midlands	2,030	2,070	2,070	2,070	2,090	2,120	2,120	2,110	2,120	2,160	2,160	2,140	2,160	2,190	2,190	2,170
West Midlands	2,390	2,440	2,430	2,440	2,460	2,500	2,510	2,490	2,510	2,550	2,540	2,530	2,540	2,580	2,580	2,550
East	2,650	2,700	2,690	2,700	2,720	2,770	2,770	2,760	2,770	2,820	2,810	2,800	2,810	2,850	2,850	2,830
London	3,860	3,950	3,970	3,980	4,030	4,100	4,110	4,110	4,130	4,190	4,170	4,180	4,190	4,240	4,240	4,230
South East	3,940	4,010	4,000	3,990	4,040	4,110	4,080	4,080	4,100	4,170	4,140	4,130	4,150	4,210	4,180	4,170
South West	2,330	2,380	2,370	2,360	2,390	2,430	2,410	2,410	2,430	2,470	2,450	2,440	2,460	2,500	2,480	2,460
Wales	1,240	1,250	1,250	1,250	1,260	1,270	1,270	1,260	1,270	1,290	1,280	1,280	1,290	1,300	1,300	1,290
Scotland	2,400	2,440	2,430	2,410	2,440	2,470	2,460	2,440	2,460	2,480	2,460	2,440	2,470	2,500	2,480	2,450
Northern Ireland	710	720	720	720	730	730	730	730	740	740	740	740	750	760	760	760

Source: HMRC PAYE RTI data

5.2 Earnings

Figure 5 compares measures of quarterly earnings based on the processed (individual-level) data against those published by HMRC.[4] As with employee numbers, earnings measures using the processed data closely resemble the published statistics. This shows that the processing applied to the data has not substantially altered the resulting statistics. The slightly higher level of earnings suggested by the processed data reflects in part the exclusion of those receiving one-off or irregular payments.

Figure 5 Mean and median quarterly earnings, 2014/15-2017/18, £

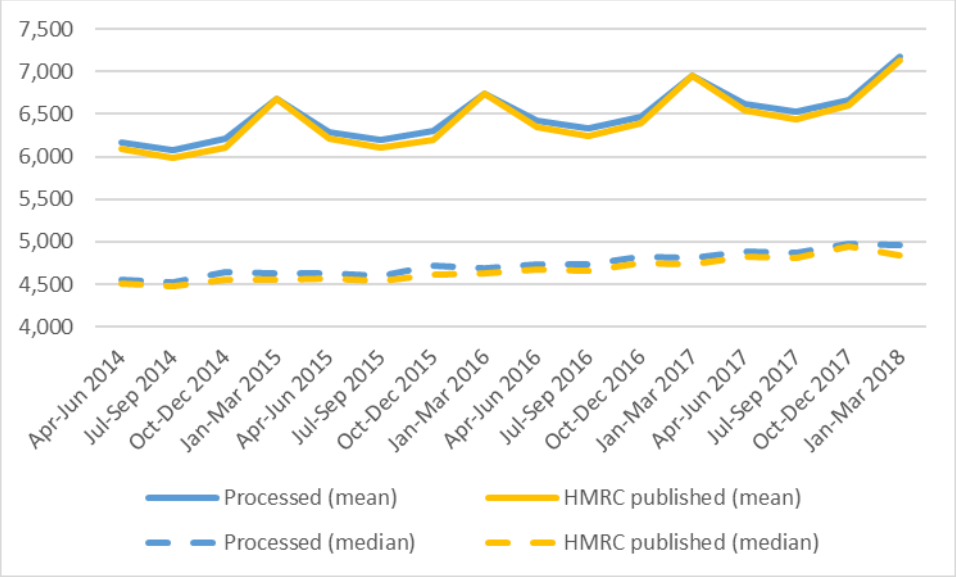
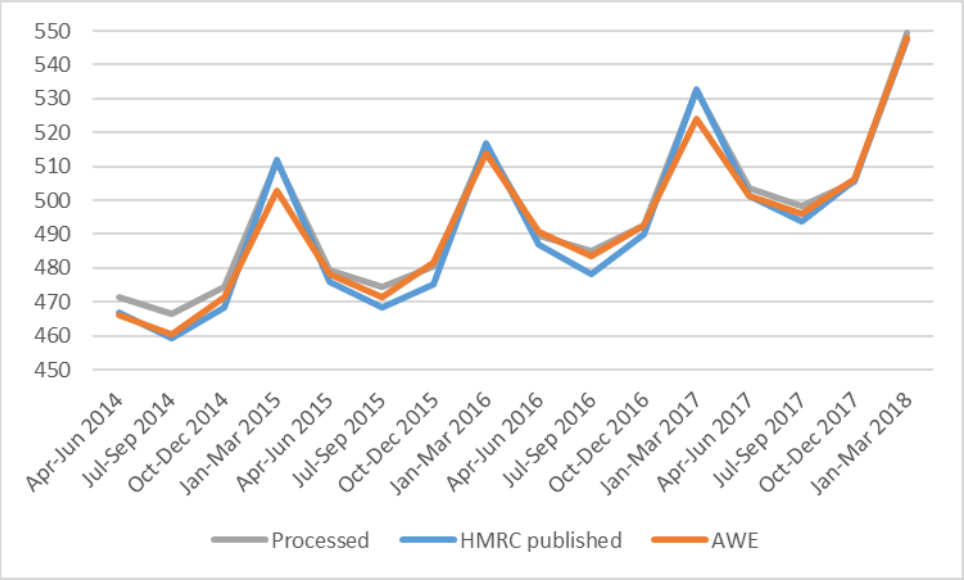


Figure 6 presents estimated of individuals' mean weekly earnings. For the processed data, this is the key earnings measure provided by the business rules described earlier on and for comparison with the published HMRC results needs only to be converted to a quarterly amount. This is done by simply averaging across the three months within a quarter. The published HMRC measures are calculated by dividing mean quarterly earnings by 13.044, the average number of weeks in a quarter. The ONS Average Weekly Earnings (AWE) series (not seasonally adjusted) is also shown. AWE is the ONS lead indicator of short-term earnings changes and is produced from the Monthly Wages and Salaries Survey (MWSS). As with the LFS, the published HMRC report includes a detailed discussion of differences between PAYE RTS and AWE measures. For convenience, the side-by-side comparison included in that report is reproduced at the end of this sub-section. However, despite these differences, the impression from Figure 6 is one of similarity across all series. After the first two quarters, the series based on processed data follows the AWE slightly more closely than the published HMRC series

[4] See Table 3 for base sizes.

Figure 6 Average weekly earnings by quarter, 2014/15-2017/18 (£)



In Figure 7, attention moves from quarterly to monthly earnings measures. This is not included in the HMRC experimental statistics, so the comparison now is just between the processed data and the AWE. Using the processed individual-level data, a distinction is drawn between mean earnings per individual and mean earnings per job (the two series differ due to multiple job-holding). Mean earnings per individual are, by construction, somewhat higher than mean earnings per job. The latter can be compared to the ONS AWE series (which is also on a jobs basis) and it is clear that the match is close.

Figure 7 Average weekly earnings by month, 2014/15-2017/18 (£)

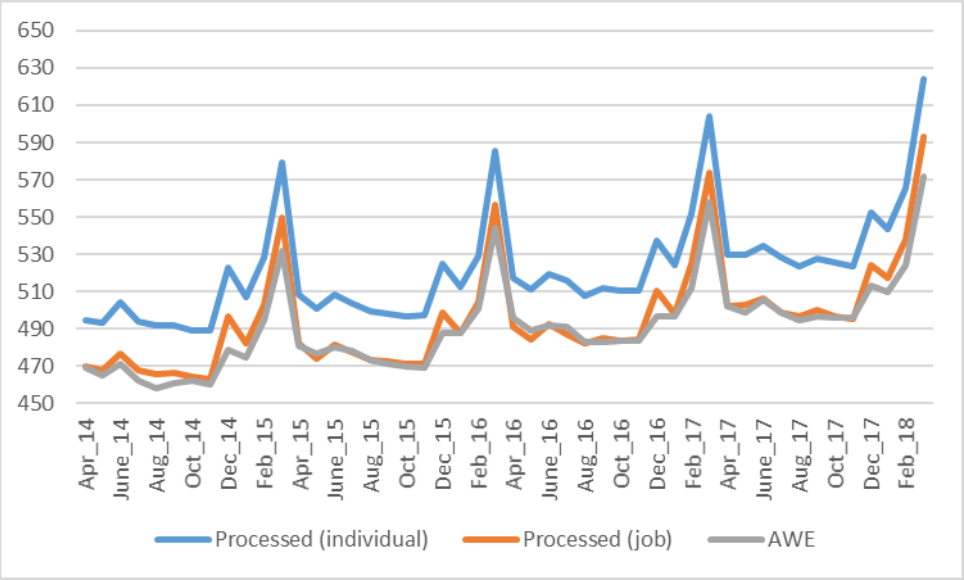


Figure 8 Side-by-side comparison of AWE and PAYE RTI data

	Average Weekly Earnings (AWE)	**Real Time Information (RTI)**
Timeliness	Published monthly. A six to seven week gap between the end of the reference period and the publication date.	Initial Publication Early 2018 as an experimental statistic. Subsequent publication schedule to be confirmed at a later date.
Employment definition	Anyone with a live employment on a PAYE scheme in the reference period	Anyone who has received pay from PAYE in the reference period
Reference period	One month	One Quarter
Average measure	Mean	Mean and Median pay from PAYE, number of individuals receiving pay from PAYE
Bonuses	AWE captures bonus payments in every month of the year, with bonuses peaking between December and April.	All bonus payments paid via PAYE are included although they can be difficult to differentiate from normal payments in the data
Inclusions	• Bonuses • Overtime • Shift premium • Allowances (weekly or monthly allowances are included in regular pay, annual allowances are included in bonus pay) • Employees on trainee or junior rates of pay • Employees whose earnings were affected by absence	RTI covers all income from PAYE so will include the following if paid via PAYE • Bonuses • Overtime • Shift premium • Allowances • Employees on trainee or junior rates of pay • Employees whose earnings were affected by absence • Payrolled Redundancy payments • Payrolled Signing on fees • Payrolled Expenses
Exclusions	• Northern Ireland • Self-employed • HM Armed Forces • Government supported trainees • Employer NI contributions • Employer contributions to pension schemes • Benefits in kind • Expenses • Arrears • Redundancy payments • Signing on fees • Stock options not paid through payroll	• Self-employed • Stock options not paid through payroll • Arrears • Employer NI contributions • Employer contributions to pension schemes • Benefits in kind
Full-time / part-time breakdown	AWE does not differentiate between full-time and part-time workers.	RTI does not differentiate between full-time and part-time workers.
Statistical adjustments	Non-response – reweighting of responses. Businesses with fewer than 20 employees are not sampled; instead they are estimated using a factor derived from ASHE	Late return grossing for open tax years.

5.3 Gross flows into and out of employment

The statistics presented so far have been cross-sectional in nature. By contrast, statistics on transitions are longitudinal in nature. Figure 9 presents estimates on gross quarterly flows into employment using the processed individual-level data and the LFS. The PAYE RTI data show a higher level of flows than suggested by the LFS. This is likely to reflect the differences between the two data sources that were noted when discussing Figure 3, which showed higher employment levels suggested by PAYE RTI data than by LFS. A strong seasonal pattern is seen with both datasets. However, this pattern appears different in either case.

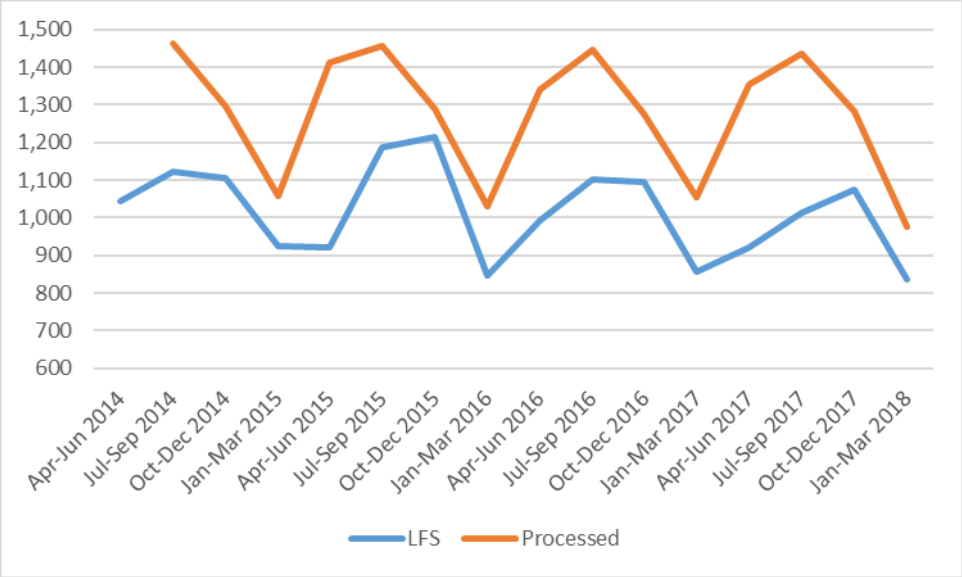


Figure 9 Gross flows into employment, 2014/15-2017/18 ('000s)

Figure 10 presents analogous results for gross quarterly employment outflows. The same comments apply although now seasonal pattern looks even more different across the two series.

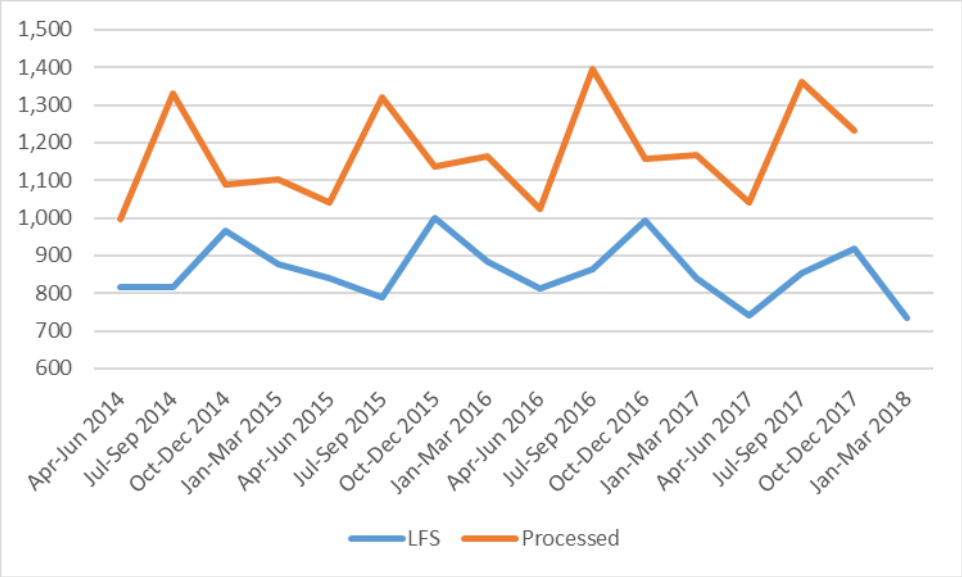
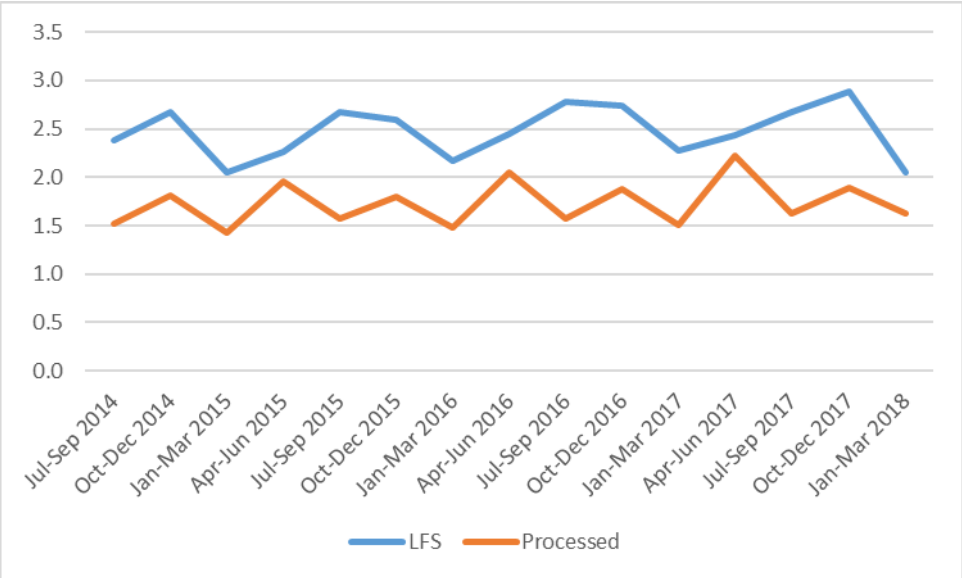


Figure 10 Gross flows out of employment, 2014/15-2017/18 ('000s)

5.4 Job-to-job flows

Figure 11 compares quarterly job flow estimates across the PAYE RTI and LFS data. In contrast to the results for gross employment flows, the LFS rates are considerably higher than those based on the PAYE RTI data. Interpreting this difference is complicated by not knowing how a job flow is defined with the LFS data (this does not appear to be documented in the published statistics). With the PAYE RTI data, we have interpreted a job-to-job flow as having arisen where an individual had a single job in the previous quarter and different (single) job in the current quarter. We recognise that this is likely to differ from the LFS definition. For instance, excluding job changes among individuals with multiple jobs will reduce the rate. For this reason, we do not expect the two series to be similar. We note though that the PAYE RTI data offer the flexibility to define a job flow in several ways, depending on what is felt to be more informative/relevant.

Figure 11 Job-to-job flow rate, 2013/14-2017/18 (%)



6 Programme of research using the processed data

The data processing described above results in two data sets:

- Jobs – a job-level dataset with the following variables:
  - nino_anon — person identifier
  - employer — employer identifier
  - mob — (Stata) month of birth
  - female — female dummy
  - sic2007 — SIC 2007 categorical variable
  - employer_size — employer size
  - wpay_XXX — average weekly pay in (Stata) month XXX
  - emp_XXX — employment (positive weekly pay) in (Stata) month XXX
  - ttwa_id — TTWA categorical variable
  - la_id — LA categorical variable
  - reg_id — region categorical variable

- People – a person-level dataset, with the following variables aggregating jobs data
  - nino_anon — person identifier
  - mob — (Stata) month of birth

- female — female dummy
- wpay_XXX — average weekly pay across all jobs in (Stata) month XXX
- emp_XXX — employment (positive weekly pay) in (Stata) month XXX
- ttwa_id — TTWA categorical variable
- la_id — LA categorical variable
- reg_id — region categorical variable

These datasets will be used to produce new statistics on labour market transitions and to estimate an econometric model of transitions. These two element are outline below.

6.1 New statistics

The processed data can potentially support a wide variety of new statistics. We would expect to agree the score of these with ONS but propose the following:

- Employment flows
  - Gross-flows into and out of employment
  - Monthly, quarterly, annual
  - Shown for three cohorts – employed in April 2014, April 2015, April 2016
  - Show breakdown by age, sex and region
- Job-job transitions
  - Need to finalise on appropriate definition of a job-to-job transition
  - Monthly, quarterly, annual
  - Shown for three cohorts – employed in April 2014, April 2015, April 2016
  - Show breakdown by age, sex, region, SIC, employer size
- Earnings mobility
  - Changes in decile/percentile ranking of average weekly earnings in a month, across all jobs for an individual
  - Incorporate non-employed (as zero earnings)
  - Monthly, quarterly, annual
  - Shown for three cohorts – employed in April 2014, April 2015, April 2016
  - Show breakdown by age, sex and region

These transitions will also be visualised as follows

- Employment flows (Sankey)
- Earnings mobility (Sankey)
- Local authority variation for statistics (map)

6.2 Econometric model

We propose to estimate an econometric model of transitions. Although this is still to be agreed and finalised, we have been focusing mainly on using the data to explore the question of how labour market tightness affects geographic mobility and pay. The approach would be as follows:

  - Estimate UK-level job matching model, allowing variation by industry.
  - Apply the results of this to predict local vacancies, V. This will be use information on local (TTWA) variations in industrial structure (calculated from the PAYE RTI data) plus publicly available information on TTWA unemployment.
  - Estimate a duration model of employment exit/entry
    - Construct sample of those employed in April 2014 who left employment at some point during the year (possible a sample)

- Competing risks – employment within same area, employment in different area
- Simultaneously model earnings of destination job
- Include labour market tightness measure, V/U, in all equations in order to assess its effect on geographic mobility and earnings mobility.