



Using Pay As You Earn Real Time Information to create timely labour market indicators: Webinar

On 17 December 2020 the RSS Official Statistics Section (OSS) delivered a webinar which was preceded by their 2020 Annual General Meeting. They invited experts from HMRC and ONS to discuss a recent major development in official labour market statistics – the use of HMRC’s real-time Pay As You Earn data to complement traditional data sources such as the Labour Force Survey and Average Weekly Earnings. This new data source has provided much-needed insight into the labour market during the COVID-19 pandemic. The webinar covered the development of the new statistics, their recent use and how it fits into the existing statistical landscape.

The meeting was chaired by the RSS OSS Secretary for 2020 Daria Gromyko. The webinar was split into two segments, the first represented by speakers who were producing the data while the second half moved on to the user perspective and Q&As.

The following is a general overview of the presentations by each speaker (Recording of the webinar is available on the RSS YouTube channel here: <https://youtu.be/lmnNz-wA31M>):

Clíodhna Taylor- Previously Head of RTI Statistics at HMRC

Clíodhna presented on methodological developments to the Pay As You Earn (PAYE) Real Time Information (RTI) statistical bulletin over the past year. She covered the challenges of adapting an administrative dataset to National Accounting principles, balancing the needs for an analytical micro-dataset with the needs for robust macro statistics, as well as developing reliable ‘flash’ estimates of payrolled employment and pay to inform our knowledge of the labour market during the COVID-19 pandemic. In addition, she spoke

about possible alternative uses of the data by statisticians and analysts. Following were key points of interest:

- **What is PAYE RTI?** Began introduction in April 2012 and fully introduced by April 2014. This is a process when employers notify HMRC with relevant information e.g. tax deductions at the time an employee is paid. All employers must register for PAYE and report under RTI where at least one employee either earns at or above the Lower Earnings Limit (£120pw in 2020/21) OR has another job. Very important to highlight that the PAYE RTI is only a dataset of employees and does not cover those that are self-employed. The RTI dataset is a dataset of payments, these payments are discrete events also the period of payment and the period of employment may differ e.g. you could be paid on 1st February for work done in January. This data gives key components to be able to calculate a continuous dataset of employment status (e.g. start date, payment dates, frequency of pay, leaving date). This then enables the calculation of the start and end of a job, as well as any gaps in employment, this is called “Calendarisation”. Further details on the methodology can be found in the following article written by Cliodhn:
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/newmethodsformonthlyearningsandemploymentestimatesfrompayasyouearnrealtimeinformationpayertidata/december2019>
- **How can we use PAYE RTI to make labour market statistics?** Cliodhna started by giving an example of someone paid every 4 weeks on Friday left work halfway between the last two payments and so the final payment is half the normal amount, so we know this persons start date and leaving date. If we naively processed this data with a crude estimate that if somebody is paid during a month then they are employed during that month and so their pay rates for that month is what their usual pay would be that month. We would get problems if we did this as summing payments in a month to calculate their monthly pay would result in spikes and defining ‘employment’ as having been paid in month would result in them being in employment in the month that follows the month they left the job, as we know they left halfway between two payments. The way we can improve on this is with Calendarisation and it has two steps, calculate a pay rate for each work period and calculate a continuous profile of employment status, so we can use Calendarisation model to improve the method. We then have a daily profile of someone with a job each day. The published data takes the daily data and how much they are getting paid and averages it over the calendar month. This data is smoother and reflects that this person is being paid a constant sum of money as per their contract.
- **How have the statistics helped us understand the labour market during the COVID-19 pandemic?** Timeliness has become more important than ever during the pandemic. The raw RTI data is “backwards looking” as we assume that jobs are paid in arrears for work done and start dates are reported alongside the first payment, and leaving dates are reported alongside the final payment. So, we do not know about someone’s employment status until the end of the relevant period. Additional complications are a small amount of data are given to HMRC after the payment date and sometimes start and end dates are not supplied or contain unlikely values. For these reasons, even after the end of a month, there is not complete data needed to calculate pay and employment for that month. However, 80-90% of data is supplied

a few days after the end of the month. Therefore, imputing the extra 10-20% of data can still provide timely statistics. Imputing administrative data in this way presents fundamentally different challenges to survey data. For example, a sample frame cannot be used to define our universe and then weight the data to match this pre-defined population total-so the data is the sample frame and the universe. A lot of thought need to go into this and was done by both ONS and HMRC in collaboration, and the methodology they derived to find someone's employment status will be by imputing two things, how many jobs have ended and how many have started.

- **What does this make possible for analysts?** Clíodhna mentioned from her perspective as an analyst there are some key features of the RTI data that point towards areas it can really add value. She mentioned analysts from HMRC and ONS could use this data for example to find out what is happening in the labour market, but academics would also find this data valuable (e.g. Longitudinal analysis).

Debra Leaker- Head of Labour Market Co-ordination at ONS

Debra Leaker discussed how the RTI statistical bulletin has fit into the ONS statistical response to COVID-19. This included a discussion of where ONS sees the advantages of RTI being, as well as where traditional data sources such as the Labour Force Survey and Average Weekly Earnings will continue to play a leading role. Following were key points of interest:

- **RTI publication development:** first joint bulletin publication by ONS and HMRC was in December 2019, this was labour market release day. In April 2020 they included flash estimates of payroll employees and that included March estimate of the payroll employees. Before that only have information from LFS and that would have been up to February, so were able to produce more timely statistics. In May 2020 they included flash estimates of median earning and again the RTI enabled them to publish more timely data when compared to their other sources. In October 2020 they published flash monthly regional data and in December 2020 they were able to publish the flash monthly industrial data.
- **COVID challenges in measuring the Labour Market:** there have been some measurement challenges due to the pandemic as the LFS survey was not able to capture information face to face and wave 1 interviews had to move to telephone to enable the data collection. Also, the Business Survey response had issues as lots of businesses were not able to work and so having an alternative data source in RTI was helpful, as were able to continue as it was administrative data. There were also methodology challenges, for example imputation and the impact of the Job Retention Scheme. There were also reporting challenges such as increased number of data sources and telling the overall story.
- **RTI vs ONS Labour Force Survey (LFS):** see table below.

RTI	LFS
All individuals receiving pay through a PAYE scheme	Sample of UK households – employment follows international definition Collected via and interviewer, information collected is self defined
Methodologies - imputation	Methodologies – weighted to population estimates Bias and sampling variability
Includes – all paid through PAYE (under 16s, those in communal establishments) Excludes – undeclared economy, schemes where no one earns above LEL	Includes – self employed, undeclared economy Excludes – under 16s, some communal establishments

- **RTI vs ONS Average Weekly Earnings (AWE):** see table below.

RTI	AWE
All employees receiving pay through a PAYE scheme	A survey of GB businesses
RTI estimates calculated on a person basis	AWE estimates are calculated on a job basis
A monthly average of daily estimates	The pay period that covers the last week of each month. Information is collected based on pay frequency
Include redundancy payments	Exclude redundancy payments

The session then moved onto the Q&A and Discussion which can be viewed in the recording of the webinar.

Overall, the webinar was very informative and the contributions from each speaker were well received, especially since this is a very busy period for them and therefore their time was much appreciated.

Written by Hira Naveed

Hira Naveed is the RSS Official Statistics Section (OSS) Secretary