

Using Administrative Payroll Data to Improve Official Labour Market Statistics in the UK

Alex Bryson^{1,2}, John Forth^{2,3}, Kyle Jones⁴ and Yohannes Lemma⁴

*Conference Paper for the Innovations in Measuring Employment Conference
17-18 September, Washington, DC*

Abstract:

In this paper we outline the features of a new and comprehensive linked employer-employee dataset (LEED) for the UK which links administrative payroll data from HM Revenue & Customs' (HMRC) Pay As You Earn (PAYE) system with information from the UK's business register, the Inter-Departmental Business Register (IDBR). This dataset enables the Office for National Statistics (ONS) to track all payrolled employees in the UK across the organisational units in which they are employed, on a monthly basis, from 2014 to 2026. We develop a methodology to align payroll reporting schemes with the statistical reporting structures used in official business surveys and use this approach to assess the coherence between administrative data and survey-based measures of the ONS's headline estimates of employment and earnings. We additionally set out a framework of options for progressively integrating administrative data into the production of labour market statistics.

Authors' affiliations:

¹ University College London

² Economic Statistics Centre of Excellence (ESCoE)

³ Bayes Business School, City St Georges, University of London

⁴ Office for National Statistics

Acknowledgements:

Any views expressed are solely those of the author(s) and cannot be taken to represent those of the Office for National Statistics (ONS). Bryson and Forth gratefully acknowledge funding from the Economic Statistics Centre of Excellence (ESCoE) Research Program.

Conflict of interest statement: The authors have no conflicts of interest to declare.

1 Introduction

The UK Office for National Statistics (ONS) is undertaking a major transformation program for its labour market statistics in response to the operational and methodological challenges that emerged during and after the COVID-19 pandemic (ONS, 2025a).

An important driver of this programme has been the downward trend in response rates across household surveys. This challenge has been further exacerbated by a long period of disinvestment in core survey operations which has impacted the ONS' ability to maintain data collection teams, replace legacy systems and re-evaluate survey designs across both household and business surveys (ONS, 2025b). Together, these issues have placed increasing pressure on the robustness and representativeness of survey-based estimates. In the case of the Labour Force Survey (LFS), which provides many of the UK's official measures of labour market activity, concerns about the quality of the underlying microdata resulted in the temporary suspension of the LFS estimates in October 2023. While dissemination of LFS estimates has since resumed, and there are signs of a gradual recovery in the quality of ONS' economic statistics across the board (Benford, 2026), the episode underscored the structural vulnerabilities associated with a continued reliance on traditional survey instruments, including surveys of businesses.

In light of this situation, the ONS has identified the need to develop more resilient and sustainable approaches for assessing, assuring, and maintaining the quality of its labour market outputs. This imperative is articulated through strategic programmes, including the Economic Statistics Plan (ONS, 2025a), the Survey Improvement and Enhancement Plan (ONS, 2025b), and the Labour Market Transformation Programme (ONS, 2025c), which collectively set out a framework for the modernisation of labour market statistics. These initiatives emphasise an increased reliance on administrative and alternative data sources, improved integration of administrative data with existing survey instruments, and strengthened quality-assurance and coherence-assessment processes.

A central component of this transformation is the strategic use of administrative payroll information collected by HM Revenue and Customs (HMRC) through employers' real-time submissions (PAYE RTI). These data provide comprehensive coverage of UK employee jobs and are available at high frequency from 2014 onwards. They capture detailed information on employment spells, earnings, tax liabilities, and social contribution or pension deductions, providing a rich and timely source of insight into labour market activity. Crucially, they are not subject to the response-rate pressures or fieldwork constraints that affect survey data.

In this paper we outline the features of a new and comprehensive linked employer-employee dataset (LEED) for the UK which links the administrative payroll data from PAYE RTI with information from the UK's business register, the Inter-Departmental Business Register (IDBR). The integration of this administrative data into ONS' labour market statistical systems will help to improve existing survey-based outputs. In particular, it will enhance the capacity of the ONS to identify potential quality issues in business surveys, support more informed sampling and weighting decisions, and provide a benchmark against which survey estimates can be validated. Such administrative data will also contribute to the production of the Labour Accounts under the updated System of National Accounts (SNA), strengthening the coherence between labour market statistics and the wider national accounts framework.

Beyond this, administrative payroll data will enable the development of novel labour market indicators that do not rely on survey data collection as well as providing a valuable resource for new and innovative research across labour market and related domains.

In the remainder of the paper, we discuss the role of LEED within the ONS' transformation program for labour market statistics. We outline the features of the LEED data and discuss the way in which LEED is currently feeding into two high-profile coherence workstreams within the ONS. Finally, we go on to outline a framework that is intended to inform wider discussions about the effective use of administrative data in the production of economic and social statistics.

Our paper contributes to debates about how administrative data can support improvements in official statistics. It does so by showing how payroll data can improve business survey-based measures of employment and earnings, and by exploring how these data can be integrated into the process of producing official employment statistics more generally.

1.1 Relevance to the US context

The situation in the UK shares a number of common points with the US experience. The PAYE RTI payroll data that form the basis of our LEED have many similarities with the administrative data collated by the US Census Bureau and used by the Bureau of Labour Statistics to generate the Quarterly Census of Employment and Wages (QCEW). The dataset we have compiled also has much in common with the Longitudinal Employer-Household Dynamics (LEHD) dataset which provides the basis for the Census Bureau's Quarterly Workforce Indicators and Job-to-Job Flows estimates.

However, the UK data infrastructure also differs from that in the US in some important respects. Principally, the UK's infrastructure is common across its four nations (England, Scotland, Wales and Northern Ireland), whereas the US relies upon administrative employment and earnings records collated via the state-level Unemployment Insurance (UI) wage reporting system. Coverage of UI can vary by state due to differences in state-level legislation (US Department of Labor, 2023) leading to inconsistencies in the types of employment activity that are measured across areas. States' filings also vary in their completeness - which we understand to be partly due to state-level variation in the cadence of data collection – leading to a need for short-term imputation and subsequent benchmark revisions. This regional heterogeneity is not the case in the UK. Finally, the processing time required to collate and harmonise data from across the 50 states delays the speed with which insights can be obtained. QCEW estimates are usually published with a lag of 6-7 months. In the UK, the ONS (working with HMRC) is able to acquire and process harmonised data from across the UK in relatively short order, publishing employment and earnings estimates within 1-2 months of the reference period (see ONS, 2026a).

The use of a single, UK-wide data collection approach has some limitations. For example, the more decentralised system in the US offers individual states opportunities to experiment with incremental enhancements (see Evermore and Raderman, 2026). Enhancements such as the collection of occupation data or more detailed hours information have been discussed in the UK but innovation has been very limited thus far. As a result of these commonalities and differences, the UK experience offers an alternative perspective to current US debates about how payroll data may be effectively leveraged for statistical purposes (Groshen et al, 2022; Yeyati et al, 2025).

2 The motivation for developing LEED

At present, official labour market statistics in the UK draw on a combination of household surveys, such as the Labour Force Survey (LFS) and its transformed successor (TLFS), business surveys such as the Short Term Employment Survey (STES) and Monthly Wages and Salaries Survey (MWSS), and some administrative sources. While each source provides valuable insights, differences in concepts, coverage, timing, and measurement can give rise to inconsistencies both within and across published outputs. Reconciling these sources, validating changes over time, and maintaining coherence across the labour market system has therefore become increasingly complex, underscoring the need for a more integrated and resilient data infrastructure.

The development of a LEED is a critical strategic response to these challenges. By linking administrative payroll data at the individual level to firm-level information from the UK's business register, LEED provides a consistent, longitudinal framework that jointly captures workers, jobs, and employers. This integrated perspective enables forms of validation, coherence checking, and analysis that are difficult or impossible when worker-based and firm-based sources are considered in isolation.

LEED has the potential to address several well-known measurement challenges associated with the LFS and business surveys capturing information on employment and earnings. Administrative payroll data provide high-frequency information on employment spells, allowing more precise measurement of short-duration jobs, overlapping employments, and job-to-job movements. These aspects of labour market dynamics are often difficult to capture accurately in surveys due to recall error, limited sample sizes for mapping specific transitions, and variation in how respondents interpret and report employment activity. By providing an administrative benchmark, LEED can therefore help strengthen the measurement of labour market dynamics and the ONS's ability to communicate labour market conditions.

Beyond improving measurement, LEED also offers a robust longitudinal benchmark for validating survey-based estimates. By comparing employment counts or earnings distributions calculated using surveys with those derived from administrative records, LEED can help identify potential misreporting, misclassification, or timing issues. For example, discrepancies between changes in employment reported in business surveys and those observed in administrative payroll data may point to reporting errors or structural business changes, while unusual wage movements identified in survey data can be assessed against administrative payroll dynamics. Together, these validation and coherence-checking exercises demonstrate how LEED can strengthen the reliability of survey outputs and improve the overall quality and coherence of labour market statistics.

The coherence work underway at ONS is instrumental in determining how LEED can most effectively support this endeavour. It will inform the appropriate positioning of LEED across a spectrum of complementary roles, including use in parallel with existing approaches to enhance validation and calibration, as well as targeted additional applications such as selective variable replacement.

Despite these advantages, the UK's LEED data infrastructure has historically been more limited than that of several comparator countries, where linked employer–employee datasets have been embedded within both statistical production and research for many years (Forth et al, 2025). Recognising this gap, the ONS has collaborated closely with academic partners through the Economic Statistics Centre of Excellence (ESCoE), to develop and evaluate a UK LEED prototype.

The development of LEED is also clearly aligned with the ONS's current Plan for Economic Statistics (ONS, 2025a), which emphasises the integration of administrative and survey data to improve labour market reporting. LEED supports this objective by enabling more systematic integration of data sources and providing an empirical basis for assessing where administrative data can enhance or partially replace survey-based estimates. In doing so, LEED contributes to the longer-term ambition of developing labour market accounts that provide a comprehensive and coherent picture of the UK labour market, using an optimal balance of survey and administrative data.

3 LEED at the ONS

The prototype Linked Employer–Employee Dataset (LEED) we present in this paper is created by linking information on workers paid via Pay as You Earn (PAYE) schemes with information on the organisations paying these employees from the UK's Inter-Departmental Business Register (IDBR).

Pay as You Earn (PAYE) is the system employers use to deduct Income Tax and National Insurance contributions (NICs) before they pay wages to their employees. Employers must submit information about employees' earnings and these deductions to the UK's tax authority (HMRC). In the 2013/14 tax year, HMRC introduced the Real Time Information (RTI) system, replacing annual reporting with a system which ensures any information about wage payments, tax and other deductions is transmitted to HMRC every time an employee is paid. This gives near real-time insight into the employment and earnings of all payrolled employees in the UK.

For businesses, the Inter-Departmental Business Register (IDBR) represents a comprehensive list of UK businesses and is used as a sampling frame for business surveys across the Office for National Statistics (ONS) and government. It covers all businesses which are VAT and/or PAYE registered and therefore only covers the registered business economy.

To facilitate the link between IDBR enterprises and PAYE schemes, we use the Longitudinal Business Database (LBD) (Lemma et al, 2023). The LBD is a set of longitudinal data spines built using consecutive snapshots of the IDBR, which contains information on the longitudinal link between enterprises and their associated PAYE schemes. Once the earnings and employment data from PAYE has been linked to the enterprise-level data from the IDBR using the LBD linkage spine, we have the ability to track all payrolled employees in the UK across the enterprise units paying them, on a monthly basis, between 2014 and 2026.

At this stage, LEED does not yet include information on self-employment due to the lag associated with individual filings under the UK's Self-Assessment (SA) system for reporting self-employed income.¹ In the future, we expect to complement the existing monthly LEED with an annual version that covers income sources beyond employee earnings.

¹ Individuals are not required to report their self-employed income to HMRC until 10 months after the end of the relevant tax year, although HMRC is currently rolling out a reporting system which will collect earnings estimates at the end of each quarter.

4 Improving existing labour market statistics

At the ONS, LEED is currently feeding into two high-profile coherence workstreams. The first looks at the UK's Short Term Employment Surveys (STES) which underpin the employee jobs private sector component of the ONS's quarterly workforce jobs estimates (ONS, 2026b). The second covers the Monthly Wages and Salaries Survey (MWSS) which is used to derive the UK's headline Average Weekly Earnings estimates (ONS, 2026c). These coherence workstreams can be viewed as attempts to assess the extent to which administrative data can be used to replicate and quality assure data collected by business surveys.

The Short-Term Employment Surveys (STES) are three statutory, quarterly surveys of private sector businesses: the Monthly Business Survey (MBS), Quarterly Business Survey (QBS), and Construction Survey (CON). The surveys ask businesses to report their number of employees (male/female/full-time/part-time) on a specified date (the first Friday after the second Thursday of March, June, September or December). Online questionnaires are sent to a sample of approximately 37,000 businesses each quarter, drawn from the Inter-Departmental Business Register (IDBR). The STES is a stratified business survey with a rotating sample design, whereby the sample is periodically refreshed while maintaining substantial overlap in responding businesses between quarters. As a result, the survey combines features of both cross-sectional and longitudinal designs and may be characterised as a rotating panel.

The combined surveys cover all sectors of the private sector economy and all industries in England, Scotland and Wales (Great Britain) with the exception of 'agriculture, forestry and fishing' and 'private households'. Data for Northern Ireland businesses are collected independently by the Northern Ireland Statistics and Research Agency (NISRA) via the Quarterly Employment Survey and are not available for use in this project. As such, our coherence work focuses on Great Britain rather than the UK as a whole.

The Monthly Wages and Salaries Survey (MWSS) is the main source of information for two key indicators of short-term earnings generated by the ONS: the Average Earnings Index and Average Weekly Earnings. The survey collects data on a firm's total gross pay bill and certain elements, such as holiday pay and bonuses, for specific groups of staff (e.g. weekly paid, monthly paid). The MWSS is a monthly survey of approximately 9,000 businesses in Great Britain, selected from the IDBR. Together, these businesses cover around 12.8 million employees. Every company with more than 1,000 employees is surveyed. Sampling is random for businesses with fewer than 1,000 employees. Companies are required to respond under the Statistics of Trade Act 1947 although, inevitably, some do not. Moreover, the MWSS does not sample businesses with fewer than 20 employees, so an adjustment factor is applied at the industry level to account for small firms in the final AWE estimates. The self-employed and government-supported trainees are also not surveyed.

Because STES and MWSS are employer surveys, coherence analysis can be undertaken by linking businesses' survey responses to the employee-level payroll records held in LEED for the same firm. For STES, LEED can identify cases where employment changes reported by businesses, or imputed during production, differ from administrative PAYE returns. For MWSS, linking to PAYE earnings data can help identify unusual wage movements in the survey and assess whether they reflect genuine payroll dynamics. Taken together, these exercises provide an evidence base for decisions about how LEED can be used to improve the coherence and reliability of survey outputs and strengthen the overall quality of labour market statistics.

5 Aligning administrative and survey data

Any analysis of the coherence between survey and administrative data at the level of the firm must consider the degree of alignment between the business units in each source. Within the UK's business register, an enterprise is described as the smallest combination of legal units which has a degree of autonomy within a wider enterprise group. These legal units are informed by administrative records from Companies House (CH), Value Added Tax (VAT) and Pay as You Earn (PAYE). An enterprise carries out one or more activities at one or more locations, known as local units.

For the purpose of statistical reporting in business surveys, enterprises are made up of one or more reporting units. These reporting units sit between an entire enterprise (the whole business) and local units (individual sites like a single shop or factory). They represent the observation unit of the firm and are used for the purpose of sampling in the majority of business surveys, including STES and MWSS. An enterprise may have one or more reporting units, depending on its complexity.

Most enterprises have one PAYE scheme, but some operate multiple schemes covering different types of staff (e.g. monthly paid and weekly paid) or different types of remuneration (e.g. base pay and bonuses). Within an enterprise, PAYE schemes should always be attached to a single enterprise (aside from a small number of pooled PAYE schemes allowed by HMRC). For enterprises with only one reporting unit (see Figure 1), their PAYE scheme(s) will align exactly with the coverage of the reporting unit, making the assessment of coherence between the survey and administrative data relatively simple. However, for more complex enterprises with multiple reporting units (see Figure 2), we must apportion individuals within PAYE schemes to specific reporting units so that administrative payroll data can be aligned with information reported in business surveys. In the specific case of Figure 2, it is necessary to allocate the employees in PAYE schemes B and C to the Reporting Units B1 and B2.

Figure 1: Simple business structure with only one reporting unit

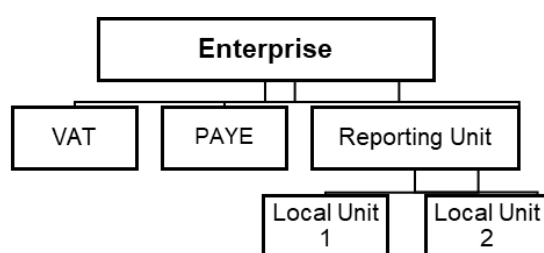
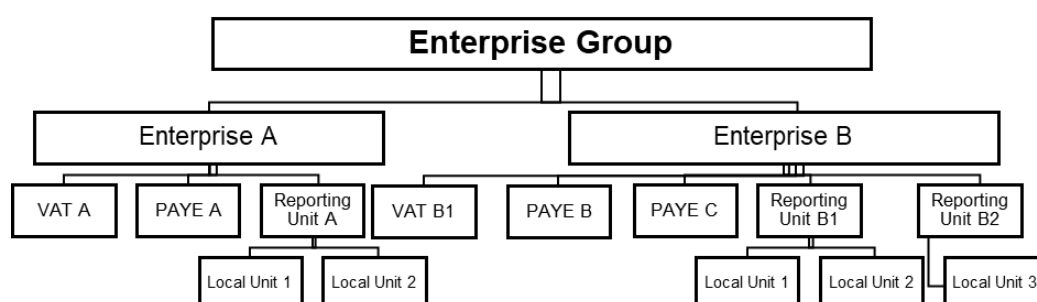


Figure 2: Complex business structure with multiple reporting units



To align administrative data on workers at the PAYE scheme level with the reporting unit structure used in survey reporting, we implement two apportioning methods. The first represents our baseline approach in which we aggregate all information on employment and earnings in LEED to the enterprise level and then apply an assumed reporting unit employment share using the latest reporting unit share information available on the IDBR. Initial coherence work, outlined below in more detail, shows that for business surveys, this baseline method of apportionment works well and closely mirrors what specific reporting units report on surveys, albeit with less accuracy in complex businesses than in the case of simple business structures.

Our more involved apportionment method looks to apportion individuals to local units (employment sites) based on the employee's residential location. This approach is necessary because, in enterprises with multiple local units, the PAYE data do not identify the local unit at which the employee is based. To achieve the apportionment, we implement a Minimum Cost Flow Assignment which assigns individuals to local units based on the Euclidean (straight line) distance between prospective sites and the employee's home address. When assigning individuals, we specify the number of employees we expect to be assigned to each site and we additionally optimise so that, at the enterprise level, the total distance travelled is minimised across all workers. We then use the local unit to reporting unit structure contained in the IDBR to assign workers to their corresponding reporting unit. This approach builds on a similar apportioning methodology developed for the construction of the US Longitudinal Employer-Household Dynamics (LEHD) dataset (see Abowd et al, 2005: 174-179). In Figure 2, our approach involves allocating the employees in PAYE schemes B and C to Local Units 1-3 in Enterprise B, and then aggregating to Reporting Units B1 and B2.

6 Coherence Results

This section presents the initial results from our coherence analysis, which evaluates the extent to which employment and earnings measures derived from LEED align with corresponding estimates collected through STES and MWSS. For each comparison, we restrict the analysis to survey respondents and link each reported survey value to its administrative-data counterpart in LEED. The population coverage of LEED means that an administrative measure can be constructed for every responding survey unit. We then compare administrative and survey-based measures of employment and earnings using the baseline business-unit apportionment approach described in Section 5.

6.1 Short Term Employment Surveys

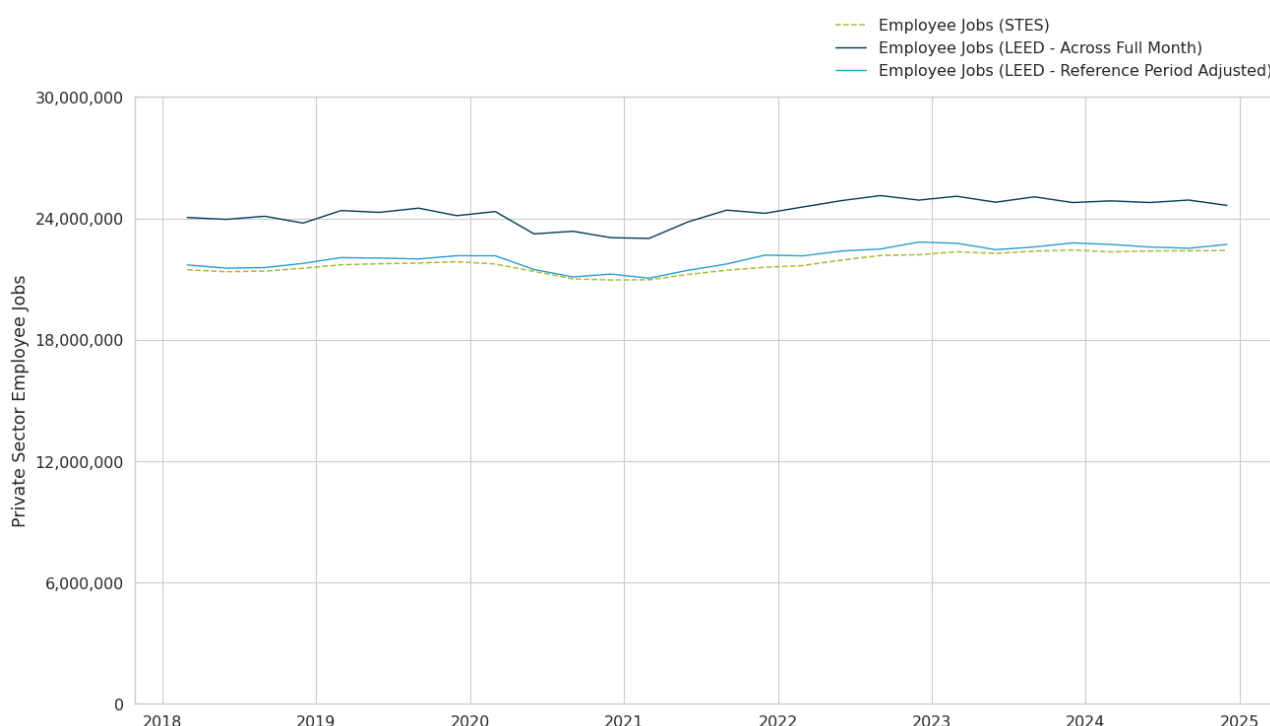
Figure 3a plots total private sector employee jobs (PS-EJ) for the linked sample of firms observed in both STES and LEED. The dark blue line shows the LEED-derived total, prior to any additional conceptual adjustment, while the dotted green line shows the corresponding total based on employment reported by firms in STES. Both estimates are weighted using the same survey weights. Holding the sample constant, the unadjusted LEED-based measure is consistently higher than the STES-based measure, although the two series evolve in a similar manner over time.

A key explanation for this level difference is the distinction between the employment concepts captured by the two sources. In the LEED calendarisation process, an employee job is recorded in a given month if a payment is made at any point for that month. By contrast, STES measures employment on a specific reference date.

As the STES reference date excludes employees who left before that date or started afterwards, or employees without guaranteed hours who did not work on the reference date, we would expect the unadjusted LEED-based measure to exceed the STES-based measure.

Using worker-level start and end dates, we align the administrative measure more closely with the STES reference-date concept by excluding LEED records for employees who left before the reference date or started after it. The resulting adjusted administrative estimate, shown by the light blue line in Figure 3a, is much closer to the STES-reported measure of PS-EJ. This provides evidence of strong agreement between the two sources once the main conceptual difference in the timing of employment measurement is accounted for.

Figure 3a: Private Sector Employee Jobs, LEED vs STES, Linked STES-LEED sample, weighted, 2018 to 2024



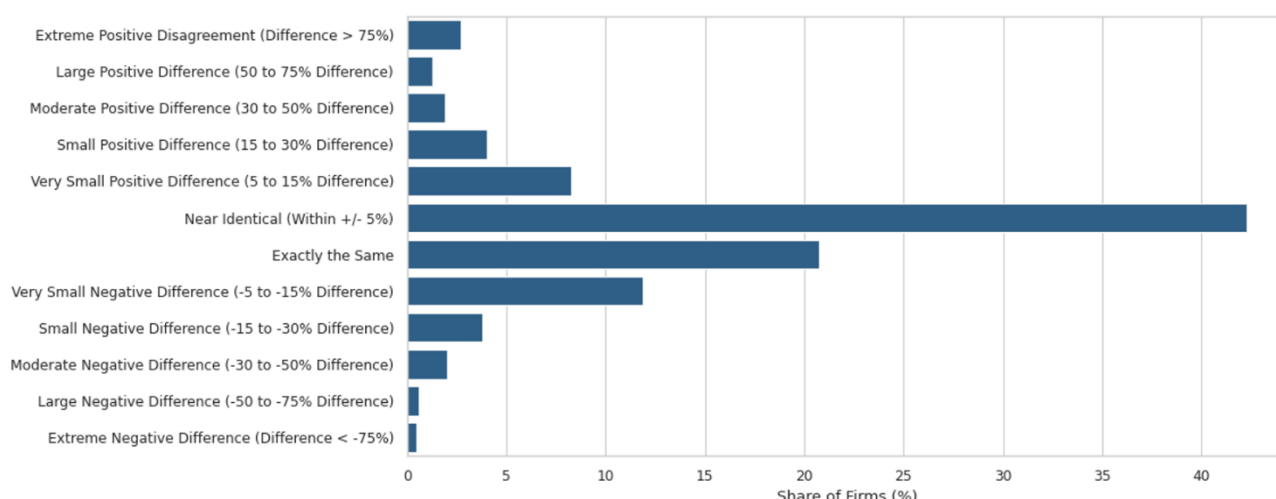
Despite this close aggregate alignment, some differences remain at the firm level. To examine these discrepancies, we calculate the percentage difference between LEED-derived and STES-reported employee jobs for each firm². Figure 3b shows the distribution of this measure for the 4 STES waves across 2024. Around one fifth of firms report an employee-jobs count in STES that exactly matches the corresponding LEED-derived estimate. A further 42% report a value within 5% of their administrative-data counterpart on this normal percentage difference basis.

More broadly, the firm-level differences between LEED-derived and STES-reported employee-job estimates are centred close to zero and are approximately symmetric. This pattern suggests that, while discrepancies exist for individual firms, there is limited evidence of material systematic divergence between the two sources.

² For a given firm, a positive difference indicates that the LEED derived measure records more employee jobs than the firm reports in STES.

When comparing firm-level differences across simple and complex enterprises, we find that the baseline apportionment method performs well. Errors for complex enterprises, where survey and administrative units do not align exactly, are only slightly larger than those observed for simple enterprises, where the reporting unit and administrative payroll structure are directly aligned. Future work in this space will model the firm-level characteristics associated with these discrepancies, with a view to informing survey weighting design and supporting more targeted response chasing during survey collection.

Figure 3b: Percentage Difference between Employee Jobs Reported in LEED vs STES, Linked STES-LEED sample, 2024



Disaggregated analyses provide similarly encouraging evidence for demographic breakdowns. LEED-derived estimates of male and female employee jobs closely reproduce the gender composition reported in STES, suggesting that the two sources provide a consistent picture of workforce demographics. Coherence is weaker for full-time and part-time employment classifications, reflecting the more limited information on working hours available within PAYE Real Time Information (RTI) data. Addressing this limitation remains an important area for further development. Planned work will explore the use of Annual Survey of Hours and Earnings (ASHE) data to calibrate full-time and part-time classifications using observed worker, employer, and earnings characteristics.

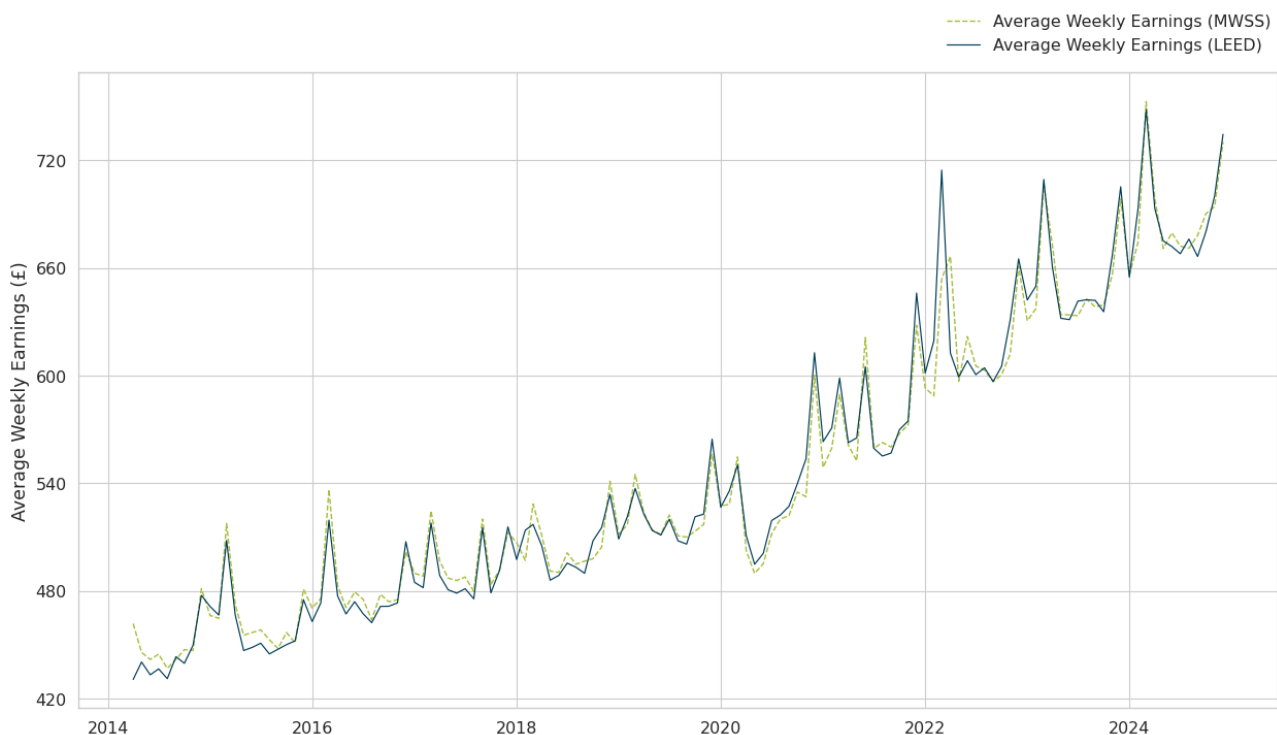
6.2 Monthly Wages and Salaries Survey

Figure 3c plots average weekly earnings (AWE) for the linked sample of firms observed in both MWSS and LEED. The LEED-derived series is constructed from total pay reported through PAYE RTI³, while the MWSS series is based on earnings reported directly by firms through the survey. Holding the sample constant, the administrative-data measure broadly reproduces the level and time profile of the survey-based AWE series, indicating a high degree of aggregate coherence between the two sources.

³ AWE in LEED is estimated by calculating a daily gross pay per employee via the calendarisation process. This is then multiplied by 7 (to reflect a full week) then aggregated to the firm level to give total weekly payroll. At the firm level, AWE is estimated as total weekly payroll divided by the number of employees employed within the month.

A key distinction between the two sources concerns the earnings concepts they capture. LEED measures an employee's total pay as recorded in administrative payroll submissions and therefore provides comprehensive high-frequency information on payrolled earnings. By contrast, MWSS is designed to support the official AWE methodology and separately collects information on both total pay and the components of remuneration such as bonus payments and arrears. As these components are not fully itemised within PAYE RTI, complete conceptual alignment with all elements of AWE remains challenging. Nevertheless, the close correspondence shown in Figure 3c suggests that RTI-based pay provides a strong basis for replicating aggregate trends in AWE on a total pay basis.

Figure 3c: Average Weekly Earnings, LEED vs MWSS, Linked MWSS-LEED sample, weighted, 2014 to 2024

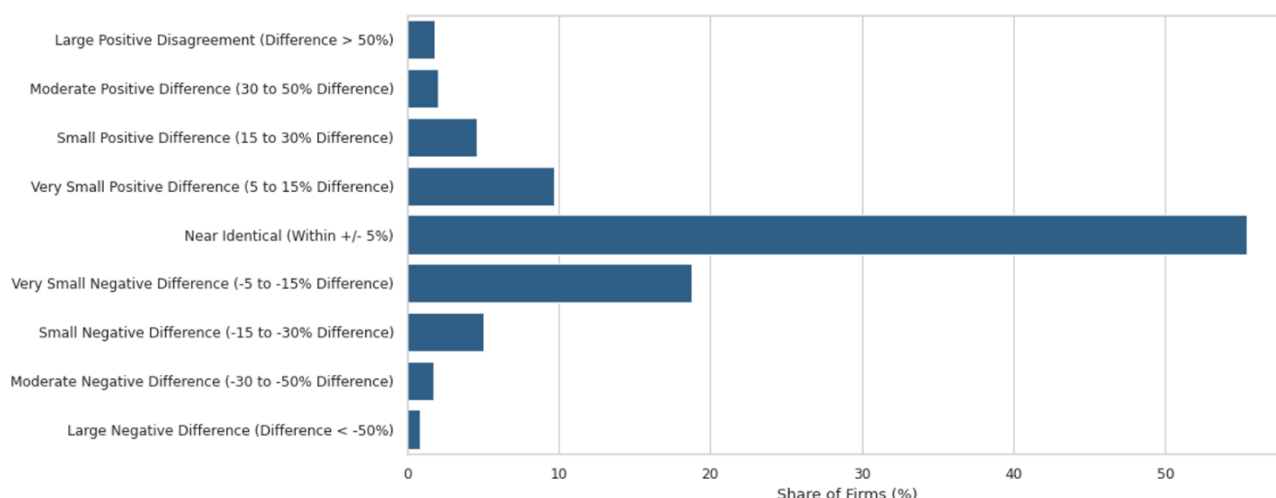


As with the STES analysis, we also examine the extent of agreement at the level of the firm. Figure 3d shows the percentage differences between LEED-derived and MWSS-reported AWE estimates across the 12 MWSS waves in 2024⁴. The distribution is centred close to zero and is approximately symmetric, with no clear evidence of systematic bias towards either the administrative or survey source. This indicates that, although discrepancies remain for individual firms, positive and negative differences largely offset one another in aggregate.

Overall, these findings provide strong early evidence that LEED earnings statistics are consistent with the established MWSS-based AWE series. In future we will be exploring the potential of LEED in the quality assurance, validation, and non-response assessment of MWSS returns. More broadly, the results highlight the value of linked administrative data as a complementary source for strengthening the production and interpretation of official earnings statistics within the UK statistical system.

⁴ A positive difference indicates that the LEED based AWE measure is greater than the AWE reported by a firm in MWSS.

Figure 3d: Percentage Difference between Average Weekly Earnings Reported in LEED vs MWSS, Linked MWSS-LEED sample, 2024



7 Operational Suitability

In addition to the conceptual coherence between administrative and survey-based measures documented above, the operational value of LEED depends on whether administrative payroll data can be processed and made available sufficiently quickly to inform the collection, validation, and production of official short-term labour market statistics.

Figures 3e and 3f compare the production timelines for STES and MWSS with the delivery schedule for LEED. Taking March 2026 as an illustrative reference period, raw PAYE RTI data are provided to ONS within the first week of the following month. Allowing approximately two weeks for ingestion, processing and indexing through the central LEED pipeline, an initial administrative-data extract is available by the middle of April 2026. As discussed in ONS (2026a), this early or “flash” extract remains subject to subsequent revision as imputed records are replaced by actual payroll submissions⁵. The second LEED delivery, which arrives one month after the first delivery, incorporates these updates and is therefore subject to substantially smaller revisions, providing the more stable version of the PAYE RTI data for statistical production purposes.

The feasibility of incorporating LEED into production differs across the two survey contexts. Because STES contributes to a quarterly indicator, there is a relatively long interval between the survey reference date and publication. This creates scope for both the flash and revised LEED extracts to be used during survey processing, quality assurance and validation. In contrast, MWSS underpins a monthly earnings indicator and therefore operates to a considerably tighter production timetable. The extent to which the revised LEED extract can be incorporated ahead of publication depends more directly on further reductions in the time required to ingest, process and index the administrative data.

⁵ Revisions can appear in later deliveries of the data, but these are very minor in comparison to updates occurring after the first revision. More information on the size of these revisions is given in the revision triangle provided as part of the ONS (2026a) publication.

If these processing times can be reduced, LEED would be well placed to meet the operational timeliness requirements for supporting the production of the ONS's short-term employment and earnings statistics. The main remaining challenges are therefore statistical and methodological rather than operational. These include managing revisions, ensuring sufficient conceptual alignment with survey measures, and assessing the stability of LEED-derived estimates in regular production.

Figure 3e: Operational Suitability of LEED relative to the STES collection and delivery timelines

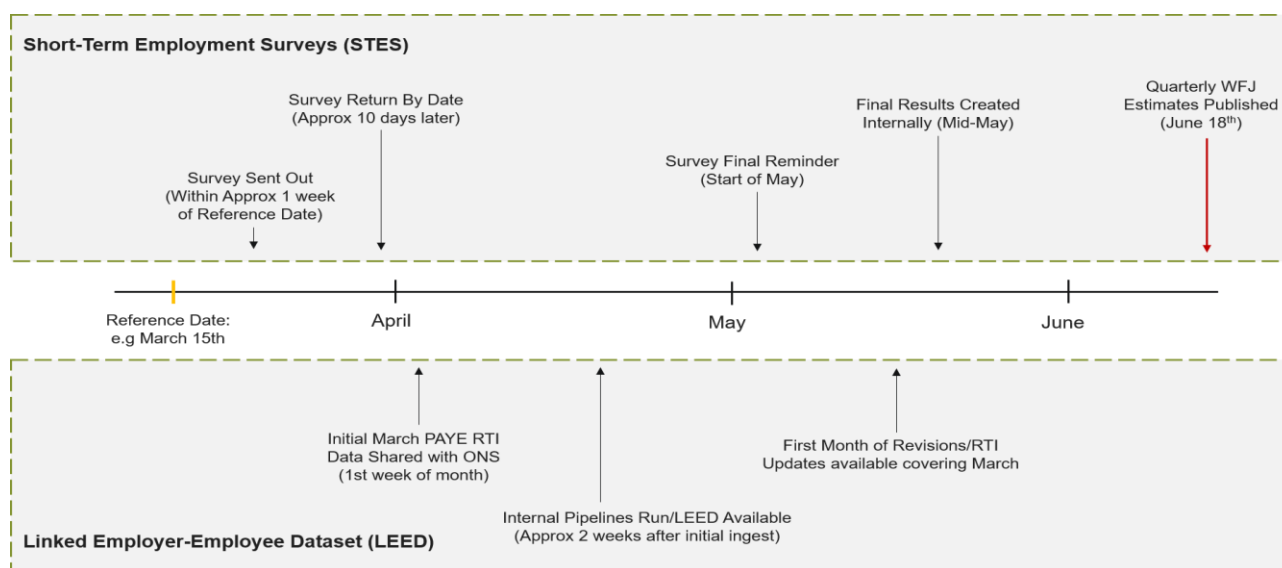
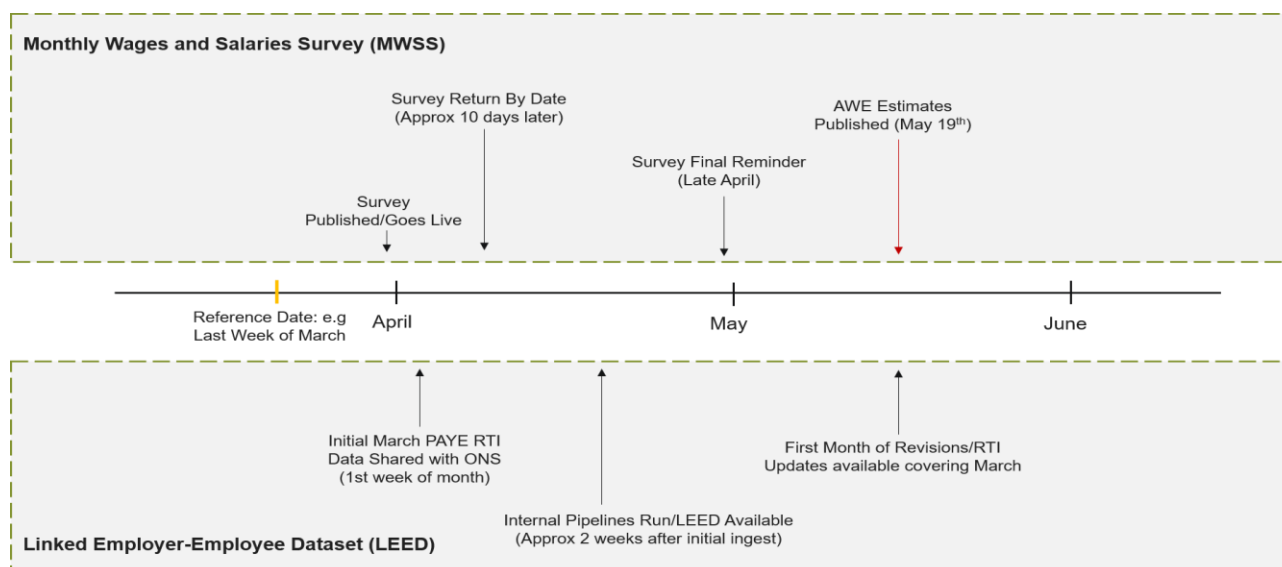


Figure 3f: Operational Suitability of LEED relative to the MWSS collection and delivery timelines



8 Implementing administrative data in the production of statistics

The increasing availability, coverage, and quality of administrative data present national statistical institutes with significant opportunities to enhance the production of labour market statistics. Realising these benefits, however, requires careful integration of administrative sources into established survey-based systems, with close attention to conceptual coherence, timeliness, data quality, and overall fitness for statistical purpose.

Building on the successful coherence exercise described above, this section sets out a spectrum of options for the practical use of the Linked Employer Employee Data (LEED) in the production of private sector employee jobs (PS-EJ) and Average Weekly Earnings (AWE) statistics (see Figure 4). These options range from maintaining the current survey only approach, through incremental and low-risk operational uses of administrative data, to more substantive integration. At each stage, coherence research would provide the empirical evidence required to support operational decision-making, ensuring that increasing reliance on administrative data strengthens rather than undermines statistical quality. We expand on each of the options below.

Figure 4: A framework of options for implementing administrative data in statistics production



8.1 Do nothing

At one end of the spectrum is the option of “doing nothing”, whereby both the Short-Term Employment Survey (STES) and the Monthly Wages and Salaries Survey (MWSS) continue to operate as standalone business surveys, with no operational use of administrative data in the production process. While this option carries no implementation risk, as it doesn’t interrupt existing production processes or statistical series, it forgoes the efficiency, resilience, and analytical gains that could be obtained from the use of LEED for coherence analysis.

8.2 Parallel production of statistics for internal quality assurance

An early and low-risk application of LEED would involve producing LEED-based PS-EJ and AWE estimates in parallel with the existing STES and MWSS production processes, accompanied by explicit internal quality assurance between the administrative and survey-based series. In this exercise, differences between the series would be systematically reviewed by the relevant production teams, with the coherence evidence used to assess whether divergences reflect genuine labour market developments, differences in statistical concepts, or potential measurement issues. Parallel production would also support wider coherence assessment across the labour market statistics system, including reconciliation with the existing PAYE RTI publication, by exploiting LEED's ability to connect individual-level payroll records to the business survey reporting structures used in STES and MWSS.

This form of parallel production would allow regular comparisons between administrative and survey aggregates to identify unexpected divergences, emerging structural changes, or issues arising during survey collection. These findings could then be incorporated into internal quality-assurance processes and, where appropriate, reflected in published commentary. The approach would therefore strengthen confidence in existing survey outputs while preserving STES and MWSS as the primary sources for official publication.

8.3 Production of new statistics to sit alongside survey estimates

An additional low-risk application of LEED would be the production of new, complementary statistics that sit alongside existing survey outputs, rather than directly intervening in their production. For STES, LEED offers the potential to publish ad-hoc PS-EJ estimates at finer industry or firm-size breakdowns where survey coverage is limited. In addition, LEED enables the production of new flow-based measures, such as hires, separations, and worker mobility between firms, which could enhance interpretation of changes in PS-EJ. For MWSS, worker-level administrative data allow more granular analysis of AWE, including decompositions of earnings growth into the contributions of job stayers, job movers, and individuals entering or exiting payrolled employment. These insights are not achievable using firm-level survey data alone and would improve the communication of changes to AWE to key stakeholders, including those involved in macroeconomic policymaking.

8.4 Operational quality assurance at firm-level in near real time

As LEED has been developed into a near-real-time data asset, a more operational use is to embed it within the survey collection process itself. Within ONS, LEED is now produced through a primary production pipeline shortly after receipt of underlying tax data, making it available internally within approximately two weeks of the reference period (see Section 7). This allows LEED outputs to be compared at the firm level with incoming STES and MWSS responses during data collection, supporting targeted response chasing and identification of potential reporting errors, and improving the outlier detection currently in place.

8.5 Improving Survey Weighting Using Population-Level Administrative Data

From a survey design perspective, LEED can also enhance both STES and MWSS by improving weighting and calibration processes. The comprehensive population-level information available in LEED allows divergence between administrative and survey measures of PS-EJ and AWE to be examined at the survey-cell level. This is particularly valuable in cells with high grossing factors and sparse survey coverage.

Coherence research helps ensure that administrative population counts are aligned with survey target populations and identifies discrepancies that arise when firms outside the sampled group behave differently from those included in the survey. Embedding these findings into the weighting strategy could reduce sampling variance while maintaining conceptual integrity and preserving the advantages of survey collection, such as the additional breakdowns not available in administrative data.

8.6 Administrative Data for Imputation and Resilience

Building on operational quality assurance, administrative data can also support imputation in cases of survey non-response. This capability is particularly important for strengthening resilience during periods of economic shock, such as those experienced during the COVID-19 pandemic where response rates to business surveys dropped. The use of administrative data for imputation relies on strong assumptions about the transportability of administrative measures into the survey framework. Coherence analysis helps to identify the variables and business types for which these assumptions are most defensible.

8.7 Targeted replacement of unsampled populations

A more substantive operational use of LEED is targeted replacement for groups of firms that are currently excluded from survey samples to reduce respondent burden. A clear case is MWSS coverage of firms with fewer than 20 employees, which are typically not sampled. Here, LEED provides a natural administrative source for extending coverage, improving the coverage and representativeness of AWE estimates without increasing burden on small businesses or costs to the ONS. It is, however, not possible to generate all AWE estimates for these firms, because LEED does not itemise bonuses, for instance.

8.8 Sample Reduction and Full Replacement

At the far end of the spectrum are options to reduce the size of MWSS and STES samples, or ultimately to fully replace them with administrative data. While these approaches offer potential efficiency gains, they also carry the greatest statistical risk. Full replacement would also inevitably lead to a reduction in the breadth of measures that could be generated, when compared with survey sources, as the administrative data lack detail in some respects.

Administrative–survey coherence research is indispensable in evaluating these options, as replacement requires administrative data to replicate not only levels and growth rates, but also the stability, revision behaviour, and interpretability of survey-based outputs. Full replacement is unlikely, because of the limits of administrative data in supplying the full range of estimates provided by surveys. As a result, surveys will continue to be valuable. However, the coherence work undertaken

as part of the LEED project will provide a valuable and robust evidence base for any future decisions on actions short of replacement.

While the immediate operational focus within ONS is on embedding LEED within the production of PS-EJ and AWE, the framework developed through this project is intended to inform wider discussions on the effective use of administrative data in the production of economic and social statistics more generally. A key component of this framework is an explicit assessment of the practical and financial implications of changes to survey collection, and how these can be balanced against improvements in quality, timeliness, and resilience. In parallel, this workstream is helping to clarify and strengthen internal governance arrangements within ONS to support future administrative data adoption.

Across this process we have engaged with a number of other national statistical institutes to learn from their experiences and approaches. Our aim is to develop a shared, evidence-based framework from which the ONS can benefit and which may also provide value to other statistical institutes undertaking similar transformation programmes.

9 Summary and next steps

In this paper we outline the features of a new and comprehensive linked employer-employee dataset (LEED) for the UK which links administrative payroll data from HM Revenue & Customs' (HMRC) Pay As You Earn (PAYE) system with information from the UK's business register, the Inter-Departmental Business Register (IDBR). This dataset tracks all payrolled employees in the UK across the organisational units in which they are employed, on a monthly basis, from 2014 to 2026.

The integration of this administrative data into ONS' labour market statistical systems has the potential to improve existing statistical outputs based on business surveys by enhancing the capacity of the ONS to identify potential survey quality issues, support more informed sampling and weighting decisions, and provide a benchmark against which survey estimates can be validated. The administrative payroll data also has the potential to enable the development of novel labour market indicators that do not rely on survey data collection. The creation of the LEED therefore marks an important step in strengthening the UK's national labour market data infrastructure.

The paper outlines the ways in which LEED is currently feeding into coherence workstreams on two ONS business surveys: the STES and the MWSS. Analysis of linked samples shows that there is a high level of coherence between survey-based estimates and estimates generated from the administrative LEED. At the same time, there are a number of statistical and methodological considerations, including the treatment of revisions, the degree of conceptual alignment with survey measures, and the stability of LEED-derived estimates in operational use, which will eventually inform the role of LEED in the production of the headline series.

The paper therefore contributes to debates about how administrative data can support improvements in official statistics. It does so by showing how payroll data can improve business survey-based measures of employment and earnings, and by exploring how these data can be integrated into the process of producing official employment statistics more generally.

The paper goes on to outline a framework illustrating the range of potential uses of administrative data in the production of economic and social statistics. We hope engagement with other statistical institutes with similar experiences will lead to the development of a shared, evidence-based framework from which we can all learn.

10 References

- Abowd, J. M., Stephens, B., Vilhuber, V. Andersson, F., McKinney K., Roemer, M. and Woodcock, S. (2005). The LEHD Infrastructure Files and the creation of the Quarterly Workforce Indicators. Chapter 5 (pp. 149-230) in T. Dunne, J. Jensen and M. Roberts (eds.) *Producer Dynamics: New Evidence from Micro-Data*, University of Chicago Press.
- Benford J (2026) "[Making progress on improving UK economic statistics](#)", ONS blog, 15th April.
- Decker R (2026) [Measuring Payroll Employment: A Note on the Current Employment Statistics](#), NBER Working Paper 34924.
- Evermore M and Raderman W (2026) [The Case for Enhanced Wage Records: Lessons from States](#), Washington, DC: Niskanen Center.
- Forth J, Bryson A and Palmou C (2025) [A Roadmap for Developing a New LEED Infrastructure for the UK](#), ESCoE Technical Paper No. 28, London: Economic Statistics Centre of Excellence.
- Groshen E, Nightingale D, Reamer A, Magdy Y and Raju M (2022) [Harnessing Employer Records for Enhanced Research, Statistics and Evaluation](#), JEDx and US Chamber of Commerce Foundation.
- Lemma Y, Lui S, Romaniuk L, Schneebacher J and Wolf N (2023) [The UK Longitudinal Business Database](#), ESCoE Technical Report No. 23, London: Economic Statistics Centre of Excellence.
- Office for National Statistics (2025a) [The Plan for Economic Statistics](#), London: Office for National Statistics.
- Office for National Statistics (2025b) [ONS Survey Improvement and Enhancement Plan for Economic Statistics](#), London: Office for National Statistics.
- Office for National Statistics (2025c) [Labour Market Transformation – Update on Progress and Plans, April 2026](#), London: Office for National Statistics.
- Office for National Statistics (2026a) [Earnings and Employment from Pay As You Earn Real Time Information, UK: March 2026](#). London: Office for National Statistics.
- Office for National Statistics (2026b) [Vacancies and Jobs in the UK: March 2026](#), London: Office for National Statistics.
- Office for National Statistics (2026c) [Average Weekly Earnings in Great Britain, March 2026](#), London: Office for National Statistics.
- US Department of Labor (2023) [Comparison of State Unemployment Insurance Laws](#), Washington, DC: Department of Labor.
- Yeyati E, Camisassa J and Seyel I (2025) [What Works for Employment Records: International Practices and Implications for the United States](#), Washington, DC: Brookings Institute.