ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Developing experimental estimates of regional skill demand

Stef Garasto, Jyldyz Djumalieva, Karlis Kanders, Rachel Wilcock and Cath Sleeman

ESCoE Discussion Paper 2021-02

March 2021

ISSN 2515-4664

DISCUSSION PAPER

Deflation of Distributional National Accounts
Stef Garasto, Jyldyz Djumalieva, Karlis Kanders, Rachel Wilcock and Cath Sleeman
ESCoE Discussion Paper No. 2021-02
March 2021

Abstract

This paper shows how novel data, in the form of online job adverts, can be used to enrich official labour market statistics. We use millions of job adverts to provide granular estimates of the vacancy stock broken down by location, occupation and skill category. To derive these estimates, we build on previous work and deploy methodologies for a) converting the flow of job adverts into a stock and b) adjusting this stock to ensure it is representative of the underlying economy. Our results benefit from the use of duration data at the level of individual vacancies. We also introduce a new iteration of Nesta's skills taxonomy. This is the first iteration to blend an expert-derived collection of skills with the skills extracted from job adverts. These methodological advances allow us to analyse which skill sets are sought by employers, how these vary across Travel To Work Areas in the UK and how skill demand evolves over time. For example, we find that there is considerable geographical variability in skill demand, with the stock varying more than five-fold across locations. At the same time, most of the demand is concentrated among three categories: "Business, law & finance", "Science, manufacturing & engineering" and "Digital". Together, these account for more than 60% of all skills demanded. The type of intelligence presented in this report could be used to support both local and national decision makers in responding to recent labour market disruptions.

Keywords: big data, labour demand, online job adverts, skills, word embeddings, machine learning

JEL classification: J23, J24, C18

Jyldyz Djumalieva, Nesta, jyldyz.djumalieva@nesta.org.uk, Stef Garasto, Nesta, stef.garasto@nesta.org.uk and Cath Sleeman, Nesta, cath.sleeman@nesta.org.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

Developing experimental estimates of regional skill demand

Stef Garasto[1], Jyldyz Djumalieva[1], Karlis Kanders[1], Rachel Wilcock[1], and Cath Sleeman[1]

[1]Nesta

March 2021

Abstract

This paper shows how novel data, in the form of online job adverts, can be used to enrich official labour market statistics. We use millions of job adverts to provide granular estimates of the vacancy stock broken down by location, occupation and skill category. To derive these estimates, we build on previous work and deploy methodologies for a) converting the flow of job adverts into a stock and b) adjusting this stock to ensure it is representative of the underlying economy. Our results benefit from the use of duration data at the level of individual vacancies. We also introduce a new iteration of Nesta's skills taxonomy. This is the first iteration to blend an expert-derived collection of skills with the skills extracted from job adverts. These methodological advances allow us to analyse which skill sets are sought by employers, how these vary across Travel To Work Areas in the UK and how skill demand evolves over time. For example, we find that there is considerable geographical variability in skill demand, with the stock varying more than five-fold across locations. At the same time, most of the demand is concentrated among three categories: "Business, law & finance", "Science, manufacturing & engineering" and "Digital". Together, these account for more than 60% of all skills demanded. The type of intelligence presented in this report could be used to support both local and national decision makers in responding to recent labour market disruptions.

Introduction

The aim of this paper is to illustrate how novel data can be combined with official statistics to enhance the quality of available labour market intelligence.

The coronavirus pandemic has highlighted how a system shock can result in very distinct impacts specific to, for example, industry [7] and location [19]. Because of this, and in line with their growing responsibilities for skills policy and funding, local governments are seeking timely and detailed indicators on local skill demand, supply and mismatch [5]. However, after the discontinuation of the UK Jobcentre Plus Vacancies series, official statistics on job vacancies are mainly available by industry and company size [25], with sparser information on vacancies disaggregated by other variables [10]. In our research, we attempt to supplement the insights from existing official statistics using novel sources of data. In this particular paper we begin by producing experimental estimates of regional skill demand and its composition by industry, occupation and skill category.

The generated insights can help decision makers to improve their understanding of employers' demand for skills and how this varies regionally and over time. Comparing local skill demand and supply may also provide more clarity on drivers of regional skill mismatches [5]. This intelligence, which is currently lacking [3, 5, 28], can inform regional policies on skill provision and retraining.

The main contributions of this paper are three-fold. First, we build on previous work to deploy a robust methodology for extracting insights on skill demand using a novel source of data : online job adverts from Textkernel (TK). The advantage of this data source is that it contains detailed information on the duration of individual vacancies. Second, we produce granular measures of skill demand disaggregated by location, occupation and skill category. These indicators are currently not captured in official statistics. Finally, we generate preliminary evidence on skill demand that could support local and national decision makers in developing response plans for labour market disruptions. This evidence, and the analysis that underpins it, could provide early-warning indicators on changes in the demand for skills or benchmarks to investigate comparisons between areas.

The generated experimental estimates of skill demand are delivered in tables showing annual average composition of demand in 2015-2019. These are broken down by geography (and further by industry), occupation and skill category. The estimates of skill demand will be released on Github[1], together with the code underlying key pieces of our methodology (such as mapping job adverts to official classifications). These resources can benefit other public sector organisations working with online job adverts.

The remainder of the paper is organised as follows. In the *Data sources* section we introduce the different datasets used in this research and the new iteration of Nesta's skills taxonomy. In the *Methods* section we describe how we convert the flow of vacancies into a stock and how we correct for biases in online job adverts. In this section, we also illustrate the process of mapping job adverts to the standard industrial classification and to categories in Nesta's skills taxonomy. In the *Results* section, we provide an overview of skill demand broken down by location, industry, occupation and skill category. We then discuss contributions of the research along with its potential applications in the *Discussion* section and conclude with suggestions for future work.

Data sources

For the purposes of the analysis, we used a range of data sources, which included official labour market statistics from the ONS, a dataset of online job adverts and a new iteration of Nesta's skills taxonomy.

ONS data sources

Vacancy Survey

We used the ONS Vacancy Survey [25] to obtain information on the number of vacancies by industry in the United Kingdom (UK). The Vacancy Survey is a monthly survey of approximately 6,000 businesses. It provides information about the total number of vacancies, for which employers are actively recruiting, by industry and company size.

The information from the Vacancy Survey was used to adjust the breakdown of online job adverts to reflect composition of vacancies by industry in the underlying UK economy.

Annual Population Survey

The Annual Population Survey [26] is a continuous household survey that provides information on social and socio-economic variables at a local level. It has a sample size of approximately 320,000 respondents and is the recommended source for employment statistics.

We used this survey to gather the number of economically active residents in Travel To Work Areas (TTWAs) in Great Britain. The number of economically active residents was used to normalise the number of vacancies that are open in a given TTWA[2]. A TTWA is a self-contained geographical area within which most people both live and work [23]. We used the TTWAs defined following the 2011 Census. Currently, there are 228 non-overlapping TTWAs covering the whole of the UK[3]. We analysed skill demand at the level of TTWAs because they are intended to approximate self-contained, and local, labour markets. One of the assumptions

[1] https://github.com/nestauk/skill_demand_report

[2] Vacancies are usually normalised by the number of employee jobs, rather than numbers of economically active residents. These two measures are related but not the same, most notably because the same person could have multiple jobs. However, the number of employee jobs is only available at the level of regions, rather than TTWAs. In future, we could approximate the number of employee jobs by adding up the number of employees to the number of people with a second job [22]. We could then align these estimates, aggregated by region, with the regional distribution of employee jobs.

[3] Employment data for TTWAs in Northern Ireland was sourced from the Northern Ireland Statistics and Research Agency and from 2011 Census [24].

behind this choice is that job seekers would be interested in data about vacancies to which they can commute. Such a geographical area is approximated by a TTWA.

Official UK industrial and occupational classifications

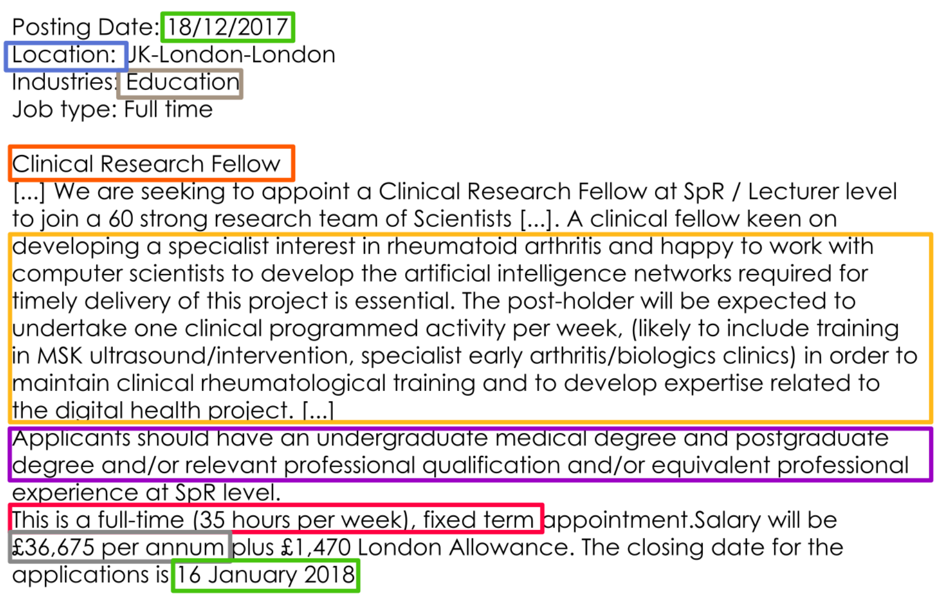
To enable the analysis of hiring activity along an industrial and occupational dimension, we mapped online job adverts to the Standard Industrial Classification 2007 (SIC 2007) and to the Standard Occupational Classification 2010 (SOC 2010), using the structure and group descriptions provided by these classifications. To analyse industry breakdown, we used broad industrial categories (e.g. 1-digit SIC, K - "Financial and Insurance Activities"). Occupational analysis was conducted up to the level of unit groups (4-digit SOC), which describe occupations at the most granular level.

Online job adverts

In this analysis, we used a dataset of online job adverts provided by Textkernel. The adverts were collected by TK by scraping active job postings from webpages on a daily basis. The total dataset contains over 40 million adverts[4], collected between March 2015 and October 2019.

Information extracted from online job adverts by TK

The dataset is a rich source of information on skill demand. As shown in Figure 1, for each job advert TK captures a number of vacancy characteristics such as job title, location, posting date, offered salary and type of contract. These characteristics are extracted by TK using a proprietary algorithm.


Posting Date: 18/12/2017
Location: UK-London-London
Industries: Education
Job type: Full time

Clinical Research Fellow
[...] We are seeking to appoint a Clinical Research Fellow at SpR / Lecturer level to join a 60 strong research team of Scientists [...]. A clinical fellow keen on developing a specialist interest in rheumatoid arthritis and happy to work with computer scientists to develop the artificial intelligence networks required for timely delivery of this project is essential. The post-holder will be expected to undertake one clinical programmed activity per week, (likely to include training in MSK ultrasound/intervention, specialist early arthritis/biologics clinics) in order to maintain clinical rheumatological training and to develop expertise related to the digital health project. [...]
Applicants should have an undergraduate medical degree and postgraduate degree and/or relevant professional qualification and/or equivalent professional experience at SpR level.
This is a full-time (35 hours per week), fixed term appointment.Salary will be £36,675 per annum plus £1,470 London Allowance. The closing date for the applications is 16 January 2018

Figure 1: Example of job advert illustrating various elements that can be extracted programmatically from text.

In addition, TK programmatically maps job adverts to official occupations (using SOC 2010) and to its own industrial classification. Since the full descriptions of adverts were provided as well, we were able to independently evaluate the quality of the occupation assignment. We have done this by manually reviewing a random selection of job adverts. Whilst there are margins for improvement, we found the SOC codes assigned by TK to be of sufficient quality to be used "as is".

[4] TK preprocesses collected adverts to identify and remove duplicates.

TK also annotates each job advert with a list of required "skills". In the context of job adverts, the term "skills" refers to a set of keywords that are deemed relevant for a specific post. In our dataset this field includes terms and phrases that describe knowledge, competences, personal characteristics, tools and certifications.

Additional indicators
For the purposes of this research, we have enriched the online job adverts with additional indicators.

To enable analysis of online job adverts by geography, we assigned each job advert to a *TTWA* based on its location coordinates and the geographic boundaries of 2011 TTWAs. In order to disaggregate online job adverts by *industry*, we have also automatically mapped job adverts to SIC 2007 (at 1-digit level). The process for deriving SIC codes is described in further detail in the *Methods* section. Finally, to later convert the flow of online job adverts to stock, we also calculated the vacancy duration as the difference (in days) between when the advert first appeared and when it was removed from online job portals.

Skills frameworks

European Skills, Competences, Qualifications and Occupations (ESCO)
ESCO is the European multilingual classification of Skills, Competences, Qualifications and Occupations. It is a rich source of information on the labour markets in Europe. One of its main resources is a freely available database that contains information on tens of thousands of standardised and occupation-specific skills. Specifically, ESCO provides descriptions for more than 13,000 skills linked to 2,942 occupations.

New iteration of Nesta's skills taxonomy
In 2018, Nesta published the first open and data-driven taxonomy of skills for the UK, using online job adverts [11]. The taxonomy was funded by the ONS as part of ESCoE (the Economic Statistics Centre of Excellence). The taxonomy represented a hierarchical grouping of skills into meaningful categories, based on co-occurrences of skills in online job adverts.

We have now produced an improved version of the skills taxonomy, which integrates the knowledge contained in the expert-derived ESCO framework with data-driven insights extracted from online job adverts. The updated skills taxonomy (henceforth referred to as combined skills taxonomy, updated skills taxonomy or skills taxonomy) was constructed using ESCO as a base and then enriching and expanding it using UK online job adverts. This approach ensures that the taxonomy still covers those occupations (such as low-paid agricultural roles) that are less frequently advertised in online job adverts.

To create the taxonomy, we first organised ESCO skills into hierarchical groups based on how often they are required together within the same ESCO occupation. This process involved representing skills as a graph, automatically clustering the skills using the Leiden community detection algorithm, and performing consensus clustering to ensure robustness of the results. Subsequently, the resulting clusters were manually profiled, evaluated and labelled based on the set of most relevant keywords among the constituent skills. We then mapped individual skills mentioned in online job adverts to the clusters in the derived taxonomy. This process is described in greater detail in Appendix 1.

The updated skills taxonomy comprises three layers. The top layer contains 15 broad clusters of skills; these are split into 76 clusters, and then further split into 201 skill clusters (henceforth also referred to as skill categories). The full structure of the taxonomy will be published in a separate report.

Methods

Online job adverts could be used to produce more timely and detailed indicators of skill demand. In principle, it is possible to build data infrastructure for analysing insights from adverts in near real-time or with a shorter lag than in existing statistical surveys. More importantly, online job adverts can enhance the granularity of estimates enabling researchers to study composition of skill demand by dimensions that are currently not reflected in official statistics. Due to these advantages, a growing number of organisations have

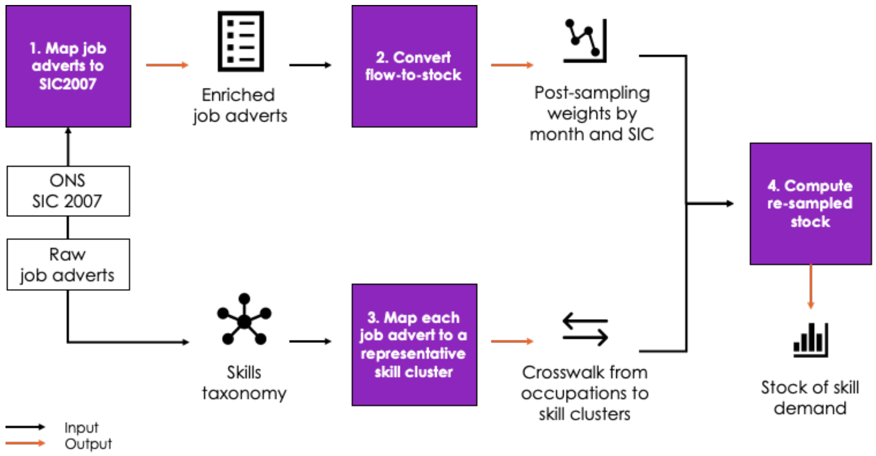


Figure 2: Summary of the process used to estimate the stock of skill demand from online job adverts.

been using online job adverts to perform economic analysis. A non-exhaustive list includes the ONS [27], Burning Glass in collaboration with the Department for Digital, Culture, Media and Sport [21], Indeed [1], also in collaboration with the Centre for Cities [19], the Institute for Employment Studies [18], the Bank of England [38, 37] and the Organisation for Economic Co-operation and Development [4].

At the same time, there are limitations to using this novel source of data. First, online job adverts may not be representative of all vacancies. There are alternatives to advertising vacancies online, which are often used in some occupations. Online postings tend to be biased toward high-skilled professional occupations, and therefore estimates of vacancy levels in the economy cannot be directly inferred from online job postings. Descriptions of roles and employer requirements also vary to a large extent in terms of completeness and language used, which makes it difficult to standardise information on required skills. Finally, there is limited historical data available. However, data availability is bound to improve with time. Furthermore, we can mitigate the remaining limitations as follows. As detailed in subsequent sections, we can align the composition of online job adverts to that in the underlying economy using data from official surveys. We can also organise tens of thousands of skills mentioned in job adverts into meaningful categories using Nesta's skills taxonomy.

The following sections provide further information on each step in the process of generating estimates of stock of skill demand (Figure 2).

Conversion from flow to stock

Mapping to SIC 2007

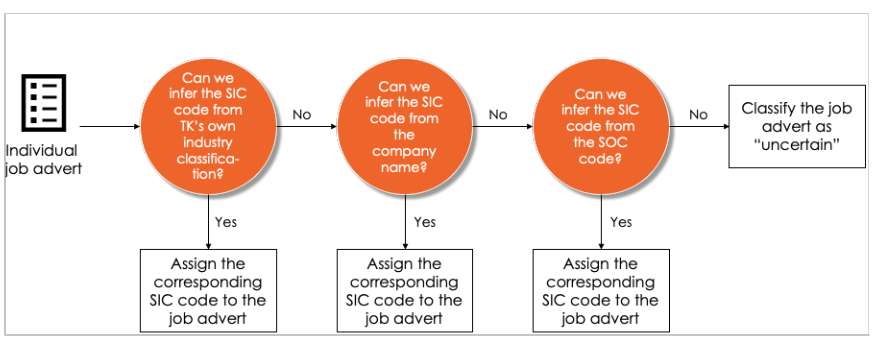


Figure 3: Summary of the process used to map online job adverts to SIC 2007.

Mapping online job adverts to SIC 2007 is an important procedure which enables us to correct for some of the misalignments between the makeup of the TK dataset and the true population of vacancies. Assigning SIC codes is also necessary for subsequent analysis of vacancies by industry.

As shown in Figure 3, we combined several strategies to programmatically assign online job adverts to SIC codes (at 1-digit level). The first approach involved using TK's own industrial classification. For each category in TK classification, we manually identified the best matching SIC codes[5]. For the most accurate results, we grouped some SIC codes together based on their relationship with TK's industrial categories (the groupings are described in Appendix 2.

The second strategy was to match names of employers (e.g. "NHS") in our dataset to companies registered with Companies House[6]. We performed automated matching of the 50 most common employers[7] to Companies House data and then evaluated their corresponding SIC codes. To improve the accuracy of matching, we also checked company names for the presence of keywords that relate to a specific industry (e.g. "police").

However, for many job adverts the name of the hiring organisation was not known, as the posting had been advertised by a recruiting company[8]. Because of this, we developed a third approach that involved building a mapping from occupations (4-digit SOC) to industries (1-digit SIC). We used semantic similarities[9] and the presence of shared known keywords in descriptions of official occupations to identify the closest matching descriptions of industrial categories. Only the first digit (e.g. "K: Financial and insurance activities") of the chosen 4-digit SIC codes was retained.

Since not all occupations are necessarily concentrated in a particular industry, we operated conservatively and only kept the strongest matches from occupations to industries[10]. The wider the spread of an occupation across industries was, the less likely we were to include that occupation in the final mapping. If none of the three strategies worked, a job advert was classified as "uncertain" (this applied to around 35% of all adverts). Overall, we prioritised having a higher level of confidence in the industry label over tagging a larger number of job adverts.

The main limitation of our approach for mapping online job adverts to industries is that it introduces a potential bias in the composition of skill demand by occupation and industry. This is because, by construction, some of the industry labels are now derived from the occupation. Therefore, in this report we do not analyse the breakdown of skill demand by industry and occupation simultaneously.

Flow to stock model

Job adverts represent a *flow* of vacancies appearing on a given day, rather than the total number of vacancies that are open at any given time. The latter is what we refer to as the *stock* of vacancies. In our analysis, we convert flow to stock for two reasons. First, in order to correct the composition of online job adverts for reasons mentioned earlier, we use data from the ONS Vacancy Survey, which measures stock. Second, in future research, we will compare indicators of skill demand and supply. Defining skill demand as a stock of vacancies and skill supply as the number of workers available to fill the openings will then enable us to assess the magnitude of potential skill mismatches [30, 13, 32].

As shown in Figure 4, to obtain the stock of online job adverts we first computed the net daily flow of postings by subtracting the number of expired job postings from the number of new job postings on that day. We

[5]This operation involved reviewing examples of job adverts within each category and was independently performed by two researchers. It was found that the two resulting maps were in agreement.

[6]Companies House is the UK's registrar of companies. Companies choose their own SIC code to list on the registrar, from a condensed list of codes. Since they can choose multiple SIC codes, we only consider the first one listed.

[7]For now, we restricted this list to 50 employers because each new organisation would need to be manually validated while adding a limited number of vacancies. However, there are ongoing efforts to grow the list of employers mapped to SIC codes.

[8]Around 55% of job adverts are classified by TK as "Staffing / Employment Agencies" (40%) or "Other / Unknown" (15%).

[9]This was computed using a word-by-word comparison between the SOC and the SIC descriptions. We calculated the final similarity score as the average cosine similarity of word embeddings (using Word2vec [20]) across all possible pairs of words (each pair being made by a word from the SOC description and a word from the SIC description).

[10]To perform the evaluation we check whether derived mappings are consistent with results produced by the first and second approaches.

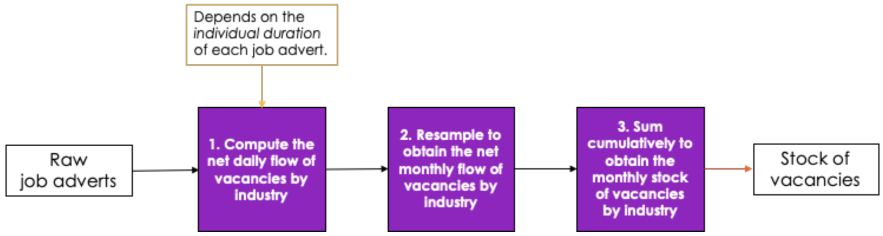


Figure 4: Summary of the steps used to transform the flow of vacancies into a stock.

then summed daily flows across all days in a month to obtain the net monthly flow. Finally, the monthly stock was estimated by adding the net monthly flow to the stock of vacancies in the previous month[11] [39]. Instead of making assumptions about the value of the stock before the period of time for which we have data available, we discarded the first two months in the time series[12].

Duration of online job adverts

In order to transform the *flow* of vacancies into the *stock* we used information about the life cycle of each job advert (that is, how long each vacancy remained open). In previous relevant studies, researchers used a fixed duration for all job adverts when calculating stock [38, 34]. This decision is justified when online job adverts are collected from a single job board or recruiter. In that case, as in a recent Bank of England paper [38], it is reasonable to assume that job adverts are posted and removed in a consistent manner after a standard period of time. This assumption may not hold true for datasets generated by aggregators of online job adverts that webscrape postings from multiple websites. A common workaround is to use the median duration of vacancies [34].

However, our analysis of duration of online job adverts shows that only 26% of postings have a duration that is within 7 days of the overall median (32 days) and that certain jobs have consistently longer (e.g. "nurse") or shorter (e.g. "customer assistant") durations. This is illustrated in Figure 5, where we show the distribution of job advert durations (in days) together with the median, and the range of job adverts that fall within 7 days of the median. We restrict the horizontal axis to be between 0 and 55 days for ease of visualisation. The observations derived from our analysis raise the question as to whether applying the median duration to all job adverts is an appropriate assumption. Being able to incorporate more detailed information on the life cycle of job adverts is therefore an improvement offered by our dataset.

In the TK dataset, the duration of each job advert was calculated as the difference (in days) between when the advert first appeared and when it was removed from online job portals. If the same job advert was posted multiple times, it was assumed to have expired when the last posting of that job had expired. Since most websites are checked on a daily basis, changes in their vacancy lists are detected with a high degree of reliability. There might be instances when job adverts are not removed from the job portal when they expire [2]; this might artificially increase the duration of those vacancies. In future iterations, we could minimise the possibility of overestimating the vacancy duration by extracting relevant information directly from the text of job adverts (e.g. "this post is no longer active").

Approximately 23% of job adverts have durations that are longer than 55 days. For these types of adverts, it is hard to determine exactly how many genuine vacancies they represent. For example, they may represent ongoing hiring needs [8] or "phantom" vacancies, that is job posts that are not withdrawn after having been filled [2, 6, 8, 31][13]. Therefore, we capped the duration at 55 days[14]. This was based on the assumption

[11]This corresponds to taking the cumulative sum of the net monthly flow.

[12]Ultimately, this did not influence the final result. This is because the whole time series changed by a constant and we subsequently adjusted the overall stock level according to the number of vacancies estimated from the ONS Vacancy Survey, as described in the following section.

[13]However, these studies have used data from France and the United States.

[14]Another option to explore in future work would be to "slice" online job posts with long durations into multiple posts, as

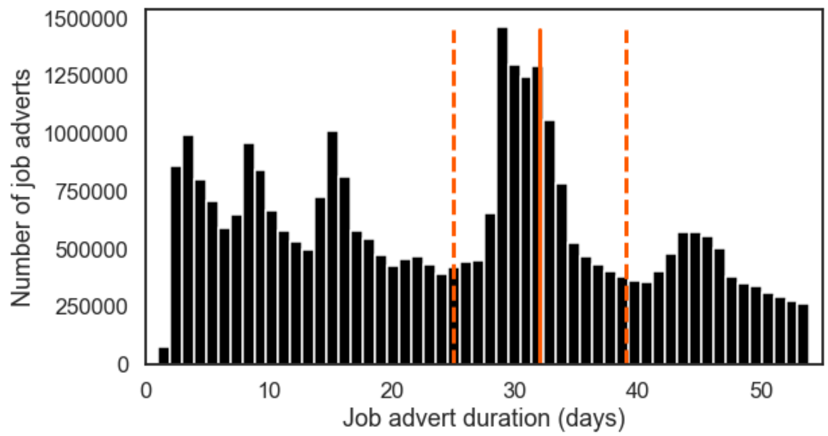


Figure 5: Histogram of the duration (in days) of job adverts (up to 55 days). The histogram shows counts of job adverts for different duration values. The solid line refers to the median duration across all job adverts; the dashed lines indicate the range of durations within 7 days from the median.

that after a certain period of time, a vacancy is less likely to receive any applicants, as some research suggests [8, 2][15]. For example, Albrecht et al. [2] report that job seekers may adopt this strategy as a reaction to the existence of phantom vacancies. Finally, very few job adverts (around 1%), for which the end date was missing, were assigned the median duration.

Adjusting the stock of skill demand by industry

For a number of reasons, such as imperfect coverage and variation in recruitment practices, the composition of online job adverts by occupation and industry is likely to be different than that in the underlying economy [38, 4]. To correct for these discrepancies, we re-weighted the stock of vacancies from online job adverts using the ONS Vacancy Survey[16]. These two datasets were aligned along their only common dimension, which is industry or SIC code.

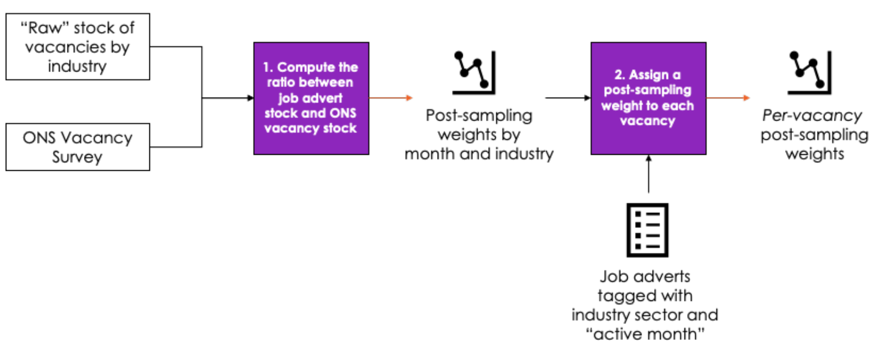


Figure 6: Overview of the process used to compute per-vacancy post-sampling weights.

done by Davis and Samaniego de la Parra [8].

[15]However, this pattern is subject to variation. For example, in the US labour market, on the one hand Davis et al. [9] report a large share of vacancies posted and *filled* within a month. On the other hand, Faberman and Kudlyak [14] found that, on the SnagAJob website, "just under 14 percent of applications are sent to newly-posted vacancies".

[16]We do not currently correct for seasonality in the stock of online vacancies, which is why we used the non-seasonally adjusted time series from the ONS Vacancy Survey. Since we report yearly averages, seasonality should have less of an effect on our results.

The methodology for adjusting the composition of skill demand by industry is outlined in Figure 6. First, we computed the ratio between the monthly stock of vacancies by industry according to the ONS Vacancy Survey, and the monthly stock of vacancies by industry according to online job adverts. This provided us with a set of "post-sampling" weights for each month and industry[17]. Our aim was to obtain a weight for each individual vacancy, as this allowed us to aggregate the weighted data along other dimensions, such as occupation and location, which are not included in the official vacancy statistics.

To derive individual vacancy weights, we matched each vacancy to an "active month". This was defined as the month in which the vacancy stays open the longest. For example, if a job advert is posted on the 14th of October 2018 and expires on the 8th of November 2018, its active month would be October 2018 (Figure 7). Each individual vacancy then inherited the post-sampling weight corresponding to its active month and to the industry to which it has been coded. Vacancies with an uncertain industry were assigned the median[18] of the monthly weights across all industries.

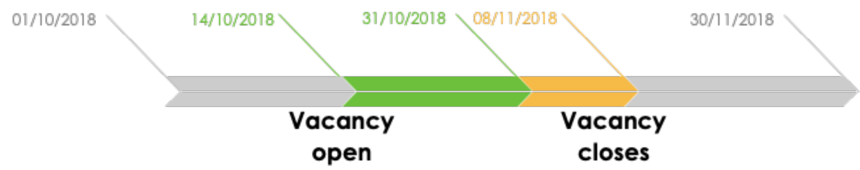


Figure 7: Example illustrating the procedure used to assign an "active month" to each individual job advert.

As mentioned above, the weights assigned to each individual vacancy can then be used to compute the stock of demand by other relevant dimensions, such as occupation or location. To do this, we aggregated the job adverts according to the variables of interest and then summed the weights of all the vacancies in each group, instead of simply counting them.

The limitations of our approach for adjusting the stock of online job adverts are twofold. First, we only correct for the composition of stock by industry and not by any other variable, such as location or occupation. The official vacancy statistics are currently not broken down by these variables. As a result, we cannot use them to control for potential biases in regional and occupational breakdown of the stock of vacancies, and instead we use the composition by these variables as observed in the online job adverts[19]. However, since the post-sampling weights are applied at the level of individual vacancies, disaggregation along these other dimensions should still produce estimates that reflect the underlying economy more accurately than the non-adjusted stock of online job adverts.

Another limitation is that we assign a weight to each vacancy based on its active month rather than taking an average of the weights across all months spanned by that vacancy. Whilst this aspect could be improved in future work, our analysis shows that the distribution of durations is similar across industries. This means that this decision is likely to affect all industries equally, and is therefore unlikely to have introduced strong biases into the analysis.

Mapping job adverts onto the skills taxonomy

In this section, we describe how we linked individual job adverts to skill clusters using the combined skills taxonomy. The linking was a necessary step for building the crosswalk from occupations to skill categories, which enabled us to disaggregate skill demand by skill category. Figure 8 shows the overall process.

[17] These weights were further adjusted to account for the existence of job adverts with an "uncertain" industry. Specifically, we computed a multiplicative monthly factor with the purpose of bringing the total stock of vacancies per month (including the "uncertain" job adverts) as close as possible to the total stock of ONS vacancies per month.

[18] The median was chosen to reduce the influence of outliers from the Mining and Quarrying sector.

[19] In the future we plan to explore the use of other data sources that can help addressing this limitation.

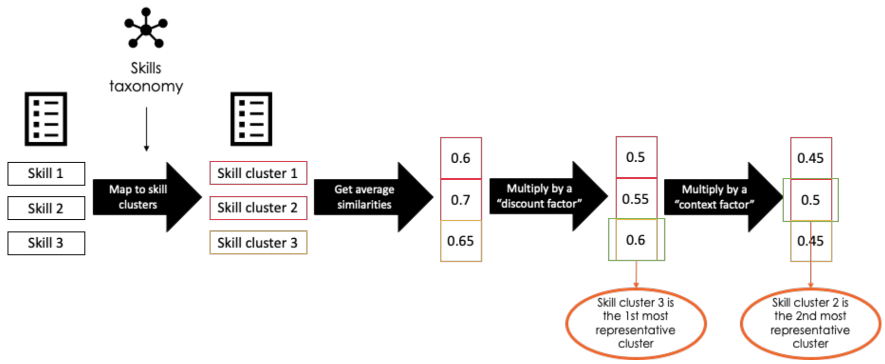


Figure 8: Overview of the process used to link individual job adverts to skill clusters.

Mapping job adverts to skill categories

First, we mapped each skill mentioned in a job advert to its respective skill cluster in Nesta's updated skills taxonomy, which gave us a list of candidate skill clusters. We then computed the average semantic similarity[20] between each candidate cluster and all the skills referenced in the job advert. This produced a "score vector" with as many entries as candidate skill clusters. We multiplied each entry in this score vector by a "discount factor", which was inversely proportional to how often a skill cluster appeared in the whole job advert dataset[21]. This was in order to prevent job adverts being disproportionately assigned to the most popular skill clusters. We then considered the cluster with the highest score as the first representative cluster.

However, we also checked whether a large proportion of skills mentioned in an advert belonged to the same higher level skill cluster. We refer to this indicator as a shared "context" and use it to identify the second representative cluster. This procedure also allowed us to take into account that a job advert might belong to more than one skill category[22]. In future work, we aim to improve our approach by expanding on the concept of multiple representative clusters. The idea would be to compute relevance weights for all candidate skill clusters, rather than limiting the choice to two clusters.

Building a crosswalk from occupations to skill categories

Finally, we produced a crosswalk from occupations to skill categories. To generate the crosswalk, we began by counting how often online job adverts were simultaneously assigned to a given occupation and to a given skill category. We did this for both representative clusters and then took the average of observed counts. These were then normalised to obtain a probability distribution from occupations to skill category, which was later used to convert the stock of vacancies by occupation into a stock of vacancies by skill category. To account for changes over time, we computed a different crosswalk for each year.

The crosswalk shows the skill categories that are often required by employers in each occupation. For example, the 2019 crosswalk for "Programmers and software development professionals" (SOC code 2136) shows that the most important skill cluster at the most granular level in the taxonomy is "Programming", associated with 64.2% of job adverts assigned to this occupation). This is followed by "Data science & data engineering" (10.3%) and "Web development" (4.8%). Among the ten most important skill clusters for this occupation we also see "Engineering", "Marketing" and "Financial management & investing", perhaps showcasing the domain knowledge that is often required by workers in this occupation.

[20] The average semantic similarity was computed as the average cosine similarity between the sentence embeddings [29] of the skills referenced in the job advert and representative sentence embeddings for the skill clusters. Specifically, the average was taken across the skills referenced in the job adverts. The representative embedding for each cluster was computed as the average sentence embedding across the skills comprising the cluster.

[21] For each skill cluster the discount factor is given by the negative logarithm of the proportion of time that skill cluster is mentioned in the job adverts, normalised to be between 0.8 and 1.2.

[22] It is possible for the two skill clusters to be the same.

Results

In the first part of the results section, we demonstrate the importance and impact of adjusting the composition of the stock of online job adverts by industry for improving the representativeness of the dataset. In the second part we provide new granular estimates of skill demand. These estimates are broken down by industry, location, occupation and skill category. Aside from the breakdown by industry, none of the other variables are currently available from official statistics on vacancies at the same level of granularity.

Representativeness of job adverts by industry

Table 1: The table illustrates the representativeness of job adverts by industry and the share of vacancies for each industry (averaged across 5 years, 2015 to 2019).

Industry*	**Ratio (pre-/post- adjustment)**	**Vacancies share (%)**
Information and communication	1.99	5.36
Construction	1.48	3.16
Manufacturing	1.46	6.72
Transportation and storage	1.26	4.35
Human health and social work activities	1.19	15.72
Financial and insurance activities	1.03	4.28
Educational and professional activities	0.98	15.18
Arts, entertainment and recreation	0.87	2.35
Personal and public services	0.81	6.19
Utilities (energy, water and waste)	0.72	0.92
Wholesale and retail trade; repair of motor vehicles and motorcycles	0.68	17.01
Accommodation and food service activities	0.66	11.12
Administrative and support service activities	0.60	6.632

* We have not included industries that are not monitored by the ONS Vacancy Survey ("Agriculture, forestry and fishing" and "Activities of households as employers") because they have not been re-weighted based on survey data. Instead, the post-stratification weights for these industries were computed by averaging the monthly weights across all other industries. Due to this, shares do not sum to 100 exactly.

This section illustrates the degree to which *online* job advertisements are representative of *all* job adverts. Specifically, Table 1 shows the industry-by-industry ratio between the distribution of the stock of vacancies before and after we re-weight the online vacancies to ensure alignment with the ONS Vacancy Survey. We use this ratio to measure the extent to which a given industry is over or under-represented in online job adverts. If the ratio is more than 1, it means that the industry is likely to be over-represented in the online job adverts dataset. The opposite is true if the ratio is less than 1.

Table 1 shows that the three most likely over-represented industries are "Information and Communication", "Construction" and "Manufacturing". These observations might increase our confidence in the estimates for occupations and skill categories that are particularly concentrated in these industries. The three most likely under-represented industries are "Administrative and support service activities", "Accommodation and food service activities", "Wholesale and retail trade; repair of motor vehicles". When we compare these findings with previous work [37], we find that "Accommodation and food service activities" is consistently found to be under-represented in online job adverts. Given that the ONS Vacancy Survey includes vacancies that are not advertised online, these results might reflect differences in the propensity of different industries towards advertising vacancies online. The industries that are under-represented might be less likely to advertise online, potentially due to the costs associated with online advertisements or the prevalence of informal recruitment practices.

Insights on industries not currently included in the Vacancy survey

One advantage of online adverts is that we can use them to estimate the stock of vacancies for "Agriculture, forestry and fishing" and "Activities of households as employers", which are currently not included within the ONS's vacancy data. Figure 9 shows the estimates of the monthly stock of vacancies for these two industries (non-seasonally adjusted). While the former industry seems to exhibit a drop in demand towards the end of 2018, the average level of demand for jobs in "Agriculture, forestry and fishing" is steadily increasing.

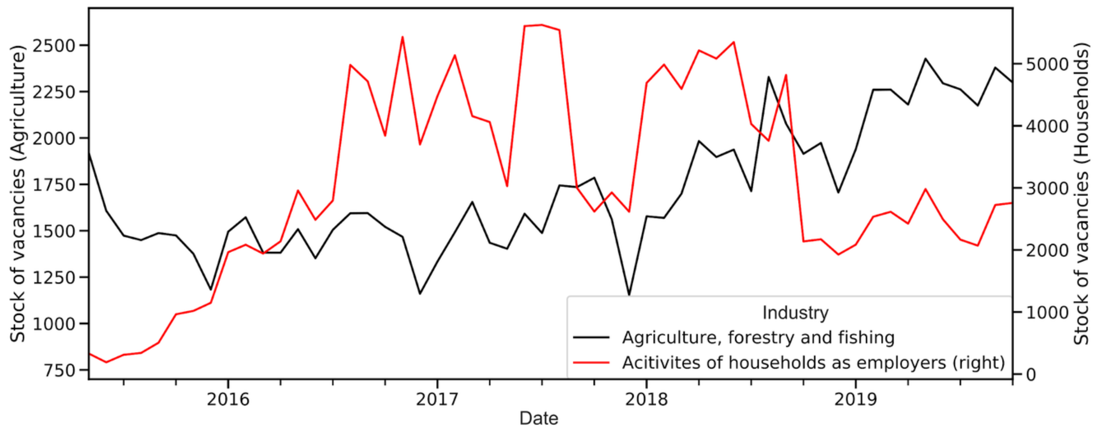


Figure 9: The monthly stock of demand (non-seasonally adjusted) for "Agriculture, forestry and fishing" and "Activities of households as employers".

A similar finding is reported in a paper by the Organisation for Economic Co-operation and Development on the representativeness of Burning Glass data. Cammeraat and Squicciarini [4] show that the number of online job postings for "Skilled agricultural, forestry and fishery workers" increased between 2012 and 2018. While the exact numbers are not directly comparable (authors use the flow rather than the stock of vacancies and study occupational composition rather than industrial), these two observations seem to support each other. Given that employment in "Agriculture, forestry and fishing" has remained stable between 2015 and 2019 [26], an increase in the stock of vacancies over the same time period might indicate a growing unmet demand for agricultural jobs (if the vacancies represent new jobs being created) or an increase in turnover rate.

Stock of skill demand by occupation

In this section we describe the breakdown of the stock of vacancies by occupation.

As shown in Table 2, the stock of vacancies appears to be dominated by "Professional occupations", "Associate professional and technical occupations" and "Managers, directors and senior officials". Together, these groups account for more than 60% of all openings. A detailed comparison of the stock of skill demand and supply is beyond the scope of this paper, and will be addressed in future research. At the same time, contrasting the share of vacancies with the share of employment can help us to detect occupations where demand and supply might be misaligned.

From the table we can see that the first three major occupational groups account for a higher share of vacancies than of employment, whilst the opposite is true for the remaining occupational groups. This finding is consistent with the results by Cammeraat and Squicciarini [4] using Burning Glass data[23]. There are several possible explanations for these findings. First, results may indicate that some occupations are still under- or over-represented in our estimates of the stock of skill demand[24]. This might be because we cannot directly adjust the stock of vacancies by occupation and instead rely on the observed occupational composition in the TK dataset. Differences between shares of vacancies and employment might also reflect

[23] A direct comparison is not possible because Cammeraat and Squicciarini [4] use a different occupational classification system (ISCO-08 instead of SOC 2010). Furthermore, they compute the flow rather than the stock of vacancies.

[24] This is what Cammeraat and Squicciarini [4] seem to suggest in their paper, together with a method to re-weight the job adverts based on the corresponding share of employment.

variation in turnover rates between occupations. If so, occupations with vacancies that are over-represented with respect to employment might generally have higher turnover rates.

Table 2: Share of vacancies and employment for major occupational groups (SOC 2010)*.

Major group	Occupation	Vacancies share (%)	Employment share (%)
1	Managers, directors and senior officials	13.74	10.79
2	Professional occupations	30.65	20.51
3	Associate professional and technical occupations	19.26	14.37
4	Administrative and secretarial occupations	8.56	10.26
5	Skilled trades occupations	6.28	10.40
6	Caring, leisure and other service occupations	5.95	9.15
7	Sales and customer service occupations	7.23	7.56
8	Process, plant and machine operatives	3.25	6.38
9	Elementary occupations	5.08	10.58

* The indicators were averaged across all years between 2015 and 2019, together with their respective share of employment [26]. Equivalent breakdowns for sub-major, minor and unit groups are presented in Appendix 3.

Finally, the lack of alignment between the share of vacancies and employment might indicate unmet demand. For example, "Professional occupations" have the largest difference, in absolute value, between their share of vacancies and employment, which might support the suggestion by Gambin et al. [15] that a number of occupations within this major group experience skill shortages. Together with evidence on growth in employment between 2015 and 2019, these findings might reflect the gradual switching of workers into professional occupations that have been growing in demand [33].

Changes in composition of skill demand by occupation over time

Table 3: Annual share of vacancies for major occupational groups (SOC 2010) and relative change between 2015 and 2019.

Major group	Occupation	2015 (%)	2016 (%)	2017 (%)	2018 (%)	2019 (%)	Relative change 2015 to 2019 (%)
1	Managers, directors and senior officials	14.63	14.85	13.58	13.13	12.51	-14.49
2	Professional occupations	30.94	31.22	30.3	30.11	30.66	-0.88
3	Associate professional and technical occupations	19.59	19.8	19.54	19.07	18.28	-6.67
4	Administrative and secretarial occupations	8.64	8.4	8.46	8.97	8.34	-3.44
5	Skilled trades occupations	6.15	6.13	6.42	6.02	6.67	8.48
6	Caring, leisure and other service occupations	4.21	5.6	6.19	6.55	7.21	71.26
7	Sales and customer service occupations	8.54	7.22	6.89	6.89	6.63	-22.36
8	Process, plant and machine operatives	3.08	2.83	3.18	3.61	3.56	15.6
9	Elementary occupations	4.23	3.96	5.45	5.66	6.13	45

Next, we investigate how the composition of vacancies by occupation changes over time. The results show that major occupational groups that are typically associated with lower skill levels (e.g. major groups 6, 8 and 9) seem to have increased their share of vacancies (Table 3). The only exception is "Sales and customer service occupations". This could reflect underlying factors such as an increase in turnover within these occupations.

Using online job adverts, we are able to produce estimates of skill demand by occupation at the highest level of granularity. As a result of this, it is possible to study changes in the share of demand for a given occupational group in greater detail. For example, we can compute the share of demand across all minor occupational groups (3-digit SOC) residing within the same major group (1-digit SOC)[25].

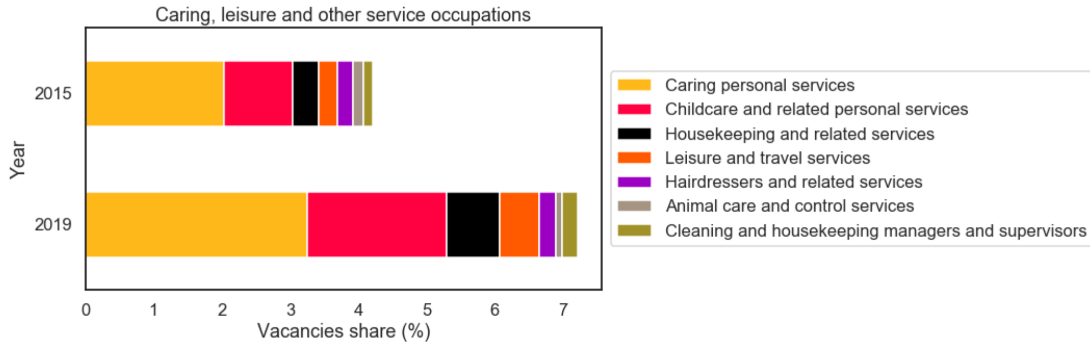


Figure 10: Share of the stock of vacancies for minor occupational groups within "Caring, leisure and other service occupations".

In Figure 10 we use the occupational group "Caring, leisure and other service occupations" as an example. Within this major group, most minor groups exhibit an increase in their share of demand at comparable rates. That is, the share of demand for 5 out of 7 minor groups[26] has a relative increase between 60% and 100%, including occupations that provide social care (e.g. "Caring personal services"). This echoes findings from previous Nesta research [12], which found an increase in the relative importance in the labour market for skills related to providing social care. Overall, the comparable growth rates (in the share of demand) for most minor groups would suggest that whilst the proportion of vacancies within this major occupational group as a whole is growing, the relative importance of its constituent occupations remains broadly stable.

Stock of skill demand by location

In this section, we analyse regional variation in the stock of vacancies across the UK. The geographical unit we use for the analysis is Travel To Work Areas (TTWAs), which have been designed to capture local labour markets - the majority of the population in a TTWA works and lives within its boundaries [23].

Figure 11 shows novel experimental estimates of the stock of vacancies by TTWA normalised by 100 economically active residents [26], averaged across all years from 2015 to 2019[27].

There is considerable variability in the relative levels of skill demand, with stock varying more than five-fold across locations. Excluding small TTWAs, the five TTWAs with the highest ratio of vacancies per economically active resident are London, Milton Keynes, Bristol, Cambridge and Cheltenham. Overall, these areas seem to concentrate around London and in the South East.

The five TTWAs with the lowest ratio of vacancies per economically active resident are Bridgend, Ballymena, Coleraine, Omagh and Strabane, and Birkenhead. It is possible that areas with a higher level of normalised skill demand is due to them having a greater concentration of industries or occupations with an increased level of hiring activity. The differences might be also related to population size [1] or levels of economic activity.

Regional variations in skill demand by industry

In this section we analyse whether there are strong geographical differences in the demand for jobs within industries. The analysis offers another example of the types of insights that we can obtain from online job adverts, which, unlike other data sources, contain highly granular information on vacancy location.

[25] Only minor groups with a share of demand of at least 0.1% are included.
[26] The two exceptions are Animal care and control services (-43%) and Housekeeping and related services (+10%).
[27] Only TTWAs with at least 40,000 economically active residents aged 16 and over are shown (the grey areas are the excluded TTWAs).

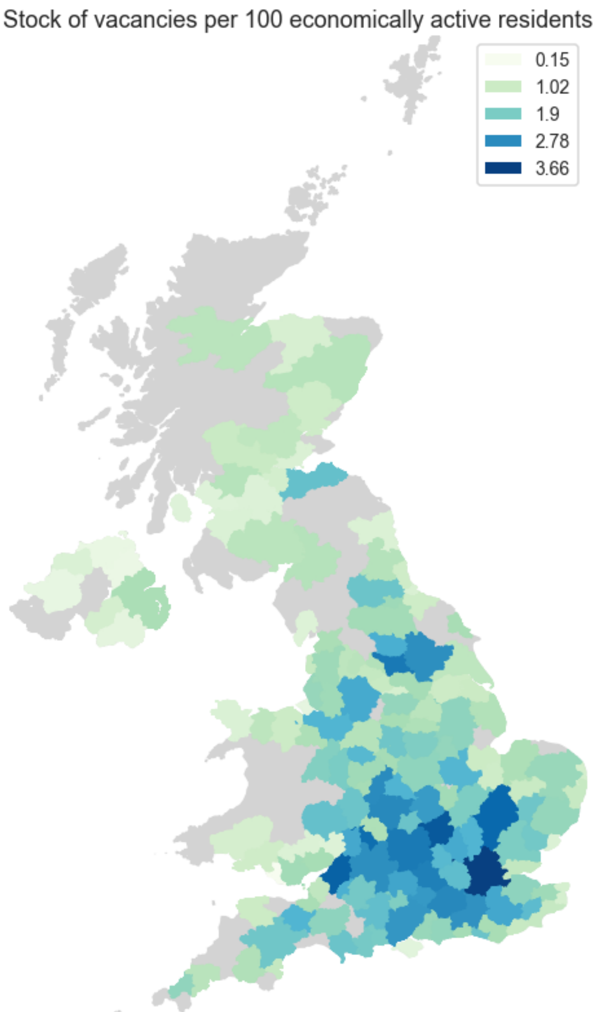


Figure 11: Stock of vacancies by TTWA normalised by 100 economically active residents.

To study variation in demand by geography and by industry, we compare location quotients across TTWAs[28] for skill demand in each industry. Location quotients measure the extent to which the demand for skills in a given industry and region differs from the same indicator for a larger region, where that larger region is the UK. To obtain location quotients we first computed the composition of demand by industry in each TTWA, and then normalised these results by the corresponding indicators for the whole of the UK. A location quotient higher than 1 indicates that industry's demand for skills is higher in a given area than its UK average. The opposite is true for location quotients lower than 1.

To measure the extent to which the skill demand in each industry varies across TTWAs, we calculated Gini indices of location quotients. The Gini index measures the inequality of a distribution and varies between 0 and 1 [16]. A higher index implies a higher degree of inequality. In our analysis, if the Gini index is particularly high for a given industry, this means that demand for vacancies in that industry varies greatly across the UK.

[28] Only TTWAs with more than 40,000 economically active residents (on average between 2015 and 2019) are included.

Table 4: Gini indices of TTWA-based location quotients by industry.

Industry*	**Gini index**
Agriculture, forestry and fishing	0.34
Utilities (energy, water and waste)	0.27
Financial and insurance activities	0.27
Transportation and storage	0.24
Information and communication	0.22
Activities of households as employers	0.19
Personal and public services	0.19
Educational and professional activities	0.17
Arts, entertainment and recreation	0.15
Accommodation and food service activities	0.14
Manufacturing	0.14
Human health and social work activities	0.12
Construction	0.09
Administrative and support service activities	0.08
Wholesale and retail trade; repair of motor vehicles and motorcycles	0.07

* Results for Mining and Quarrying are omitted as fewer than 100 online job adverts were assigned to this industry.

Table 4 shows that across all industries in the UK economy, Gini index values vary between 0.07 and 0.34. The five-fold difference in indices implies that demand for some industries (i.e. low Gini index) is substantially more uniform across regions than others (i.e. high Gini index).

To better illustrate the intuition behind Gini indices, Figure 12 contrasts the spread of vacancies across TTWAs for two industries. From the figure we can see that location quotients for "Information and communication" span a larger range (visualised as the space between the two red-dotted lines) than those for "Administrative and support service activities". This means that the former industry is more concentrated in certain areas than the latter.

Industries with the most and least uniform demand for skills across TTWAs

The three industries that have the lowest regional variation in skill demand are: "Wholesale and retail trade; repair of motor vehicles and motorcycle", "Administrative and support service activities" and "Construction". These industries are generally considered to be part of the "everyday economy", and are likely to have more uniform demand across areas, which is reflected in lower Gini indices. In contrast, the three industries where vacancies are concentrated in particular regions are "Agriculture, forestry and fishing", "Utilities (energy, water and waste)" and "Financial and insurance activities". It is possible that these industries are less evenly distributed because companies within them are more constrained to specific locations, or because they might benefit the most from clustering around other similar companies.

Hotspots for industry skill demand

Ballymena, Spalding and Craigavon are among the areas with an above average demand for jobs in "Agriculture, forestry and fishing". Likewise, the TTWAs with the highest concentration of vacancies from "Financial and insurance activities" include Edinburgh, Hereford and Bournemouth. Edinburgh is a major financial centre in the UK [36] - for example, the Royal Bank of Scotland's headquarters are located there. London also appears among the top ten TTWAs with a high level of vacancies in the financial industry. Its relatively lower ranking might be explained by the fact that in London the financial industry is *competing* with other industries, like "Information and communication", that also have a particularly high concentration.

Insights into the industrial composition of skill demand within a given area

Location quotients can also be used to assess the importance of industries to a local economy. For example, Table 5 shows the location quotients for all industries in two particular TTWAs, London and Cambridge. The table suggests that London has a higher concentration of vacancies in "Information and communication",

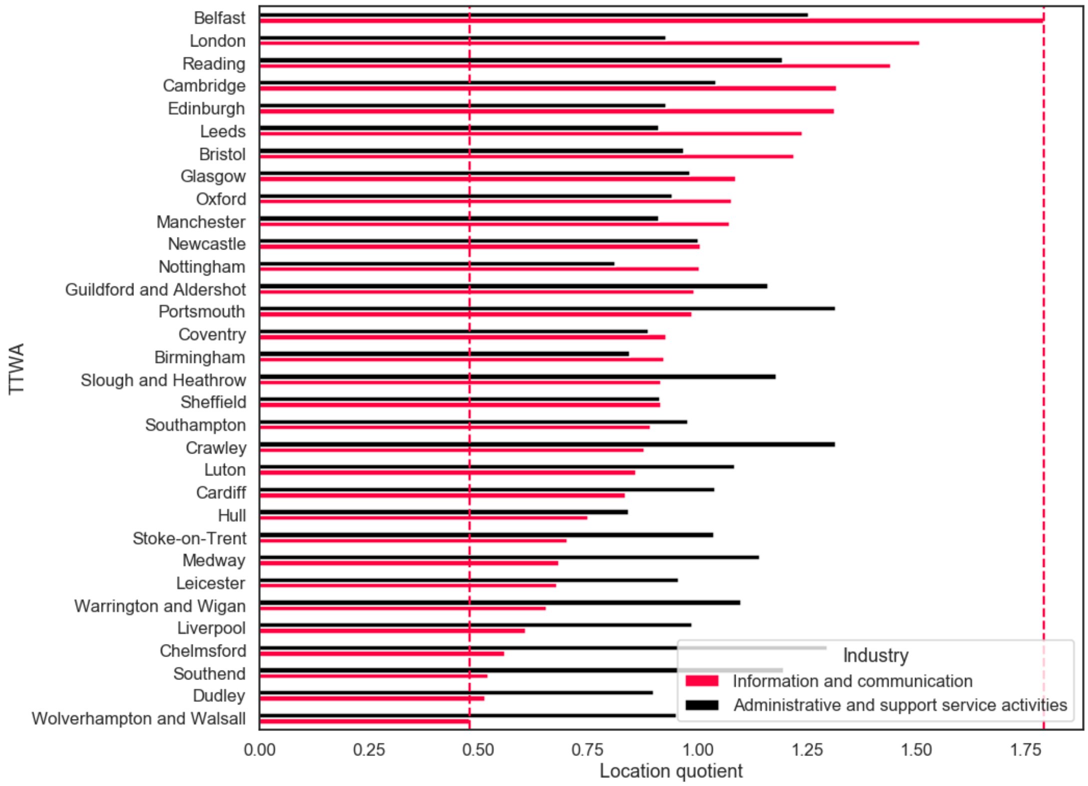


Figure 12: Distribution of location quotients for "Information and communication" and "Administrative and support service activities". For clarity of visualisation we only show TTWAs with at least 250,000 economically active residents.

"Financial and insurance activities" and "Arts, entertainment and recreation" than the rest of the UK, while the opposite is true for industries like "Transportation and storage", "Utilities (energy, water and waste)" and "Agriculture, forestry and fishing".

The three industries that are more concentrated in Cambridge than in the rest of the UK are "Information and communication", "Manufacturing" and "Educational and professional activities". These findings are consistent with a recent Tech Nation report showing that both London and Cambridge are in the top 20 European cities for tech investment [35]. Cambridge is also the home to a leading university and a hub for advanced manufacturing.

Thus far, we have only considered the stock of demand averaged between 2015 and 2019. In the future, we can also analyse how it varies over time and investigate whether some industries have become more or less concentrated on an annual or quarterly basis. Doing so could help us to detect changes in demand in a more timely manner.

Stock of skill demand by skill category

This section provides insights on the types of skills that employers have asked for most frequently in online job adverts. Figure 13 shows the composition of skill demand in the UK by top level skill categories, between 2015 and 2019. Most of the demand is concentrated among the following three categories: "Business, law & finance", "Science, manufacturing & engineering" and "Digital". Together, they account for more than 60% of the share of skill demand. The composition of skill demand also seems to remain stable across time,

Table 5: Location quotients by industry for two TTWAs: London and Cambridge.*

Industry	**London location quotient**	**Cambridge location quotient**
Information and communication	1.51	1.31
Financial and insurance activities	1.49	0.77
Arts, entertainment and recreation	1.27	0.7
Educational and professional activities	1.2	1.22
Activities of households as employers	1.16	0.87
Personal and public services	1.08	0.72
Accommodation and food service activities	0.98	1.06
Wholesale and retail trade; repair of motor vehicles and motorcycles	0.93	0.91
Construction	0.92	1.06
Administrative and support service activities	0.92	1.04
Human health and social work activities	0.82	0.86
Manufacturing	0.77	1.27
Agriculture, forestry and fishing	0.75	0.94
Utilities (energy, water and waste)	0.68	0.7
Transportation and storage	0.57	0.94

* The table is ranked by London local quotients in decreasing order.

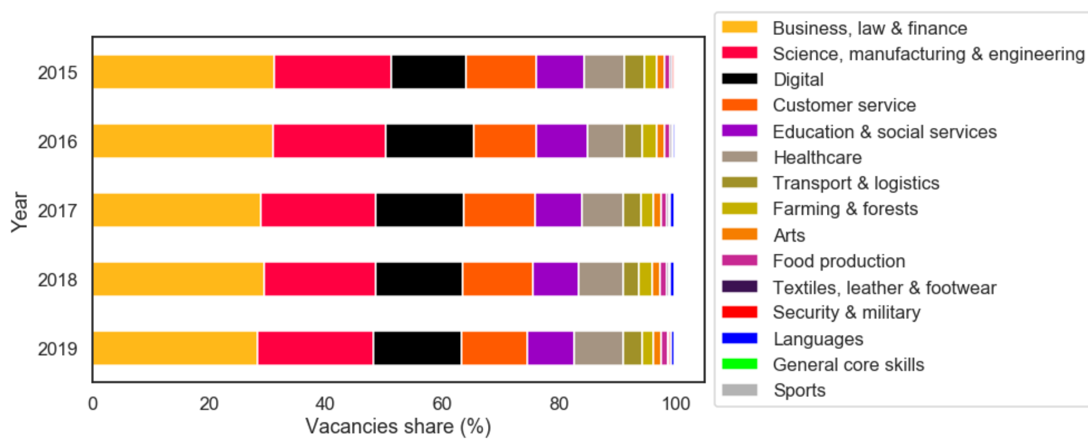


Figure 13: Composition of skill demand by top level skill clusters, between 2015 and 2019.

with some minor variations. For example, the demand for "Business, law & finance" and "Customer service" skills slightly decreased between 2015 and 2019 (-8.9% and -7.1% respectively), whilst the demand for skills in "Digital" and "Healthcare" slightly increased (+18.5% and +21.4%, respectively). The increase in the proportion of demand for IT and health-related skill categories is consistent with findings from McKinsey, where IT and Healthcare are amongst the industries with the highest predicted growth rate in Europe through to 2030 [33].

We can also analyse the breakdown of skill demand at higher levels of granularity. Figure 14 shows the proportion of skill demand for each second level skill category between 2015 and 2019. For example, we can see that "Software development" is the largest skill cluster within the broad "Digital" skill category, followed by "Data science & data engineering". This is consistent with the findings outlined in a report by Burning Glass commissioned by the Department for Digital, Culture, Media and Sport [21].

We only consider those clusters with an average share of at least 1%. Each skill cluster is labelled according to its corresponding first level category, using the same schema as Figure 13 (for example, "Financial services" belongs to "Business, law & finance" and, as such, is presented in yellow).

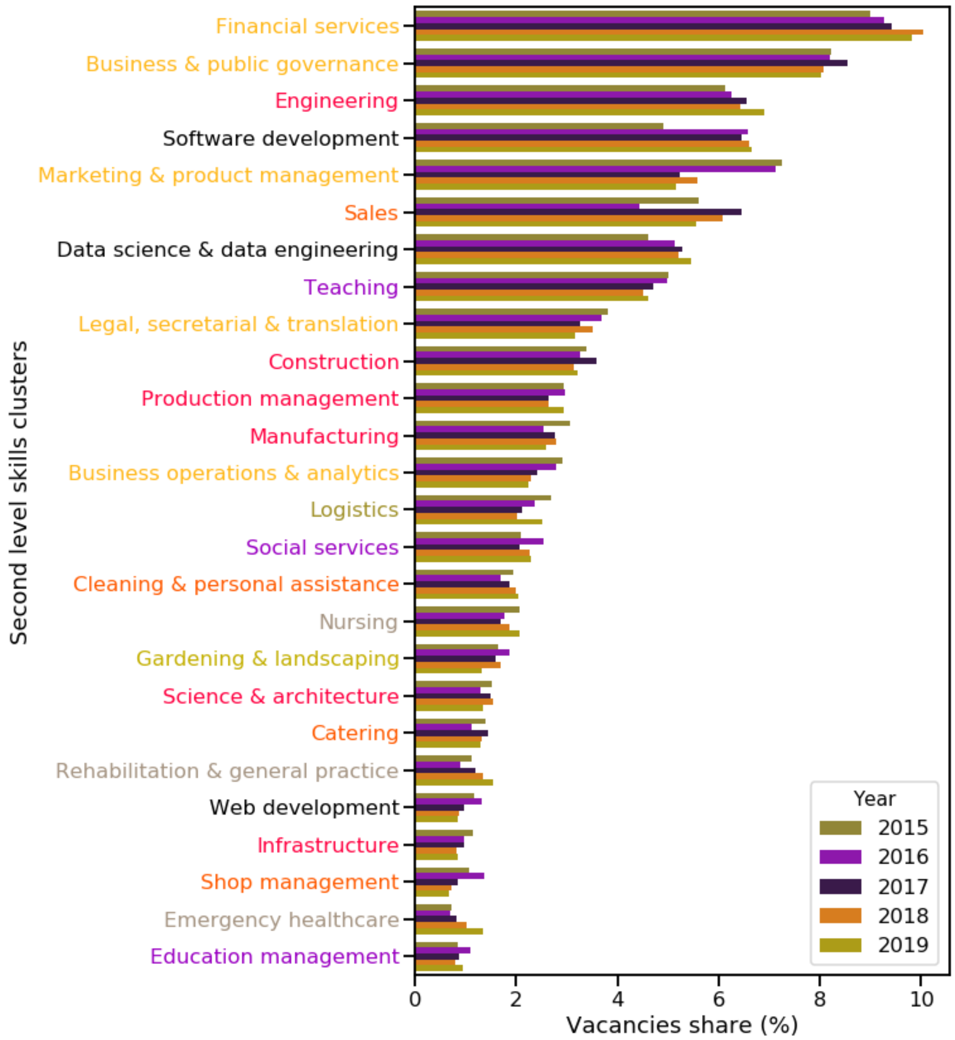


Figure 14: Composition of skill demand by second level skill clusters, between 2015 and 2019.

The share of demand for more granular skill categories within the same top level skill category do not always change in the same direction. For example, within "Business, law & finance", "Financial services" is increasing its share of skill demand whilst "Marketing & product management" and "Legal, secretarial & translation" are decreasing. In contrast, all skill clusters within the "Healthcare" category seem to be stable or increasing.

Overall, the skill clusters with the largest average increase in share of demand over the observed time period include "Emergency healthcare", "Rehabilitation & general practice" and "Software development". At the opposite end, the clusters with the largest reductions include "Shop management", "Marketing & product management", and "Infrastructure". There are several factors that could explain these findings. For example, shifts in the composition of skill demand might be driven by changes in the employment levels within particular industries (e.g. Healthcare, both private and public sectors) or occupations. Alternatively, they could indicate changes in the skill mix within occupations, or changes in recruitment patterns (i.e. increase in turnover). Further research, both quantitative and qualitative, would be needed to disentangle these factors.

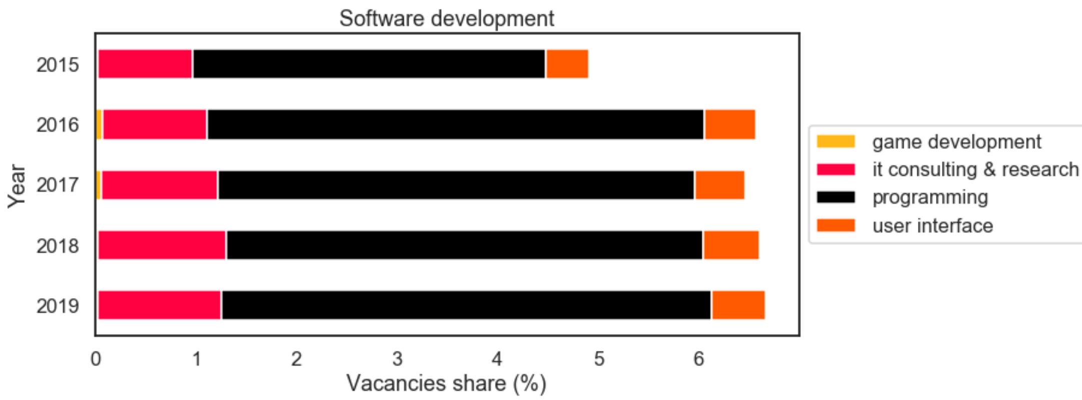


Figure 15: Yearly share of vacancies for the "Software development" skill cluster, between 2015 and 2019.

At the same time, to improve our understanding of what is driving some of the changes in the composition of skill demand, we can analyse how the makeup of more granular skill clusters varies within selected second level skill clusters. For example, Figure 15 shows changes in the share of demand between 2015 and 2019 for the "Software development" cluster[29]. From the figure, we can see that the "Programming" sub-cluster is both the most requested skill set and the one with the largest proportional increase. "Programming" includes skills like "agile software development", "software engineering" and various programming languages (e.g. "Java", "Python").

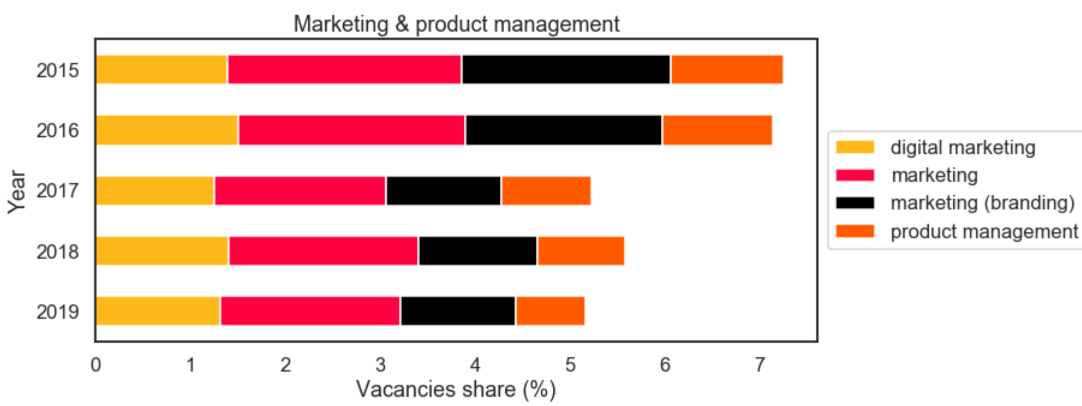


Figure 16: Yearly share of vacancies for the "Marketing & product management" skill cluster, between 2015 and 2019.

In another example, shown in Figure 16, "Digital marketing" is the sub-cluster within "Marketing and product management" with the smallest decrease in demand (5.3% decrease from 2015 to 2019, as opposed to 35% decline on average for all other sub-clusters). This means that the relative importance of this skill category within the "Marketing and product management" category increased over time. "Digital marketing" contains skills like "social media", "digital marketing", "content management", "search engine optimization",

[29]To generate this figure, we first computed the share of demand across all level 3 skill clusters and then selected the ones from "Software development".

"google analytics", whilst "Marketing" contains skills such as "marketing", "pricing strategies", "sales process", "performance indicator" and "market trend". A report by Burning Glass found that skills related to digital marketing are one of the fastest growing skill sets [21]. Whilst further research is needed to more accurately measure increase in demand in absolute terms, both reports highlight the importance of digital marketing skills in the current labour market.

Regional variation in skill demand by skill category

In this subsection, we investigate regional variations in demand for skill categories. The findings complement results on composition of demand by industry described earlier. This is because a given skill category might be in demand across different industries, and similarly, an industry might require workers with a mix of different skills.

Similar to the analysis of regional variation within industries, we computed the Gini index of the location quotients across TTWAs for skill clusters in the second level of the updated skills taxonomy[30]. We obtained Gini indices that vary between 0.04 and 0.16. Generally, a higher Gini index indicates that there are greater differences in the relative demand for a corresponding skill cluster across various TTWAs.

The five skill clusters with a more homogeneous share of demand across the UK (i.e. low Gini index) are: "Business & public governance", "Legal, secretarial & translation", "Shop management", "Sales" and "Business operations & analytics". In contrast, we observe the greatest variation in demand for skills in "Teaching", "Web development", "Catering", "Nursing" and "Software development".

A higher level of variation in demand for digital skills echoes the findings by Burning Glass on the presence of regional hotspots for this skill category [21]. Edinburgh, London, Reading, Belfast and Cambridge are among the TTWAs that appear to have a higher than average share of demand for "Software development" and "Web development" (as well as for other similar digital clusters).

Some of the other results shown are less intuitive. Specifically, we would expect demand for skills in "Teaching", "Catering" and "Nursing" to be much more uniform across TTWAs. At the same time, these findings are broadly in line with those obtained previously when analysing the makeup of employment by skill category [12][31]. One possible explanation is that the observed results could be caused by skill shortages. That is, certain areas may have an increased turnover of vacancies that require these skills because they are hard to fill.

It may also be possible that since there is stable demand for teaching and nursing skills (that is, hospitals and schools are needed everywhere), their *relative share of demand* might be subject to greater variation depending on the strength of the local economy and demand for other skill categories, such as "Digital" and "Business, law & finance". This causes the large ranges of local quotients that we see and, in turn, the higher Gini indices.

Discussion

In this paper we present granular estimates of the *stock of skill demand* , broken down by industry, location, occupation and, importantly, skill category.

Improved measurement of skill demand

The first part of this paper builds on previous work to present a rigorous methodological framework for extracting insights from job adverts. The intention is to contribute to the development of better infrastructure and measurements for labour market intelligence that could be used by others in the field. Firstly, we highlight that whilst some studies focus more on the number of job openings per day [1], converting the

[30] The Gini index is computed only across those TTWAs with more than 40,000 economically active residents (on average between 2015 and 2019). Only skill clusters with a share of demand of at least 1%, on average across years, are included in the following analysis.

[31] "Catering" is the most notable exception and more research is needed to understand the drivers behind this result.

flow into stock is an essential step towards being able to compare the stock of demand with the stock of supply [30, 13, 32]. Secondly, we discuss the importance of sourcing detailed information on the life cycle of job adverts when turning the flow of vacancies into a stock. When this information is not available, the solution is typically to use a median duration for all job adverts [34]. However, our analysis shows that job adverts have a wide range of durations, so reducing these to a single number would constitute an oversimplification. As such, this emphasises the importance of ensuring that the duration field is populated when scraping job adverts from web portals.

We also highlight how re-weighting the job adverts based on survey data increases their representativeness of the underlying economy [38, 39]. This procedure allows us to show which industries are under-represented and over-represented in our dataset of online job adverts. Such information can be of value to researchers analysing the unweighted stock of online vacancies. Finally, we introduce a new iteration of Nesta's skills taxonomy, which aims to integrate expert-derived and data-driven information on skills. More details on this taxonomy will be released in a future report.

Granular measures of skill demand

In the second part of the paper, we provide an overview of the granular measures of skill demand produced in the research, disaggregated by location, occupation and skill category. Going further than existing survey data on vacancies, we are able to show the regional variation in the stock of skill demand, further broken down by industry. This is a particularly important consideration in light of the significant regional variations in the UK economy [17, 19]. In our analysis, we find that demand for jobs in some industries (e.g. "Agriculture, forestry and fishing") is more geographically concentrated than in others (e.g. "Wholesale and retail trade; repair of motor vehicles and motorcycle"). This suggests that for job seekers, deciding where to live could be of varied importance, according to their industry of interest.

We also show the breakdown of skill demand by occupation, at multiple levels of granularity, from major to unit groups. This could be helpful for understanding hiring patterns for types of jobs, rather than simply types of industries. In the future the analyses described in this report could be conducted with higher frequency enabling us to detect changes in demand in a timely manner.

Finally, this paper shows estimates of the stock of skill demand broken down by skill category. This breakdown enables the analysis of hiring activity in terms of "bundles of skills" that are often used together in the same job. These estimates are made possible by mapping vacancies onto Nesta's skills taxonomy. In turn, this mapping helps us to generate a crosswalk between occupations and skill categories, which can be used in a multitude of ways. In the immediate future we will apply it to the analysis of skill supply statistics, so that they can also be broken down by skill category. More broadly, this mapping can provide information that is currently lacking about in-demand skill sets across the board and for specific occupations, which could then inform better career advice and course design [3, 5].

The type of intelligence that we showcase in this report could be used to support local and national decision makers in developing response plans for labour market disruptions. For example, it could assist in providing early-warning indicators on how demand for skills is changing, nationally and locally. We identify skill sets that have increased (e.g. "Software development") or decreased (e.g. "Marketing & product management") in their relative importance on a yearly basis. A similar methodology can be used to track changes in skill demand over shorter time scales. Finally, our analysis of the regional variation in skill demand (both by industry and skill category) can also provide insights into local benchmarks that could benefit governmental bodies like Skills Advisory Panels (SAPs). Indeed, research suggests that there are challenges hindering SAPs' efforts in collecting and analysing data at a local level [3].

Conclusion

In this research we used a robust methodology to extract experimental estimates of economic indicators from online job adverts. Building on previous work [38, 39], we first converted the flow of online job adverts into a stock of vacancies using detailed information on the life cycle of job adverts. Then, we used survey data to correct for biases in the industrial composition of the stock of vacancies. This debiasing step relied on our ability to map online job adverts to SIC 2007.

Using this methodology, we derived granular measures of hiring activities in the form of experimental estimates of the *stock of skill demand*. By leveraging online job adverts we were able to break down these estimates along various dimensions, such as industry, location and, most crucially, skill category. The latter relies upon the application of Nesta's skills taxonomy. We used a new iteration of the taxonomy that incorporates both data-driven and expert-derived skills. Furthermore, the online job adverts dataset we used has the advantage of including detailed information on the duration of individual vacancies, which improves the accuracy of our estimates.

We then demonstrated how these indicators of skill demand can be used to analyse which skill sets are sought by employers, how these vary across TTWAs in the UK and how skill demand evolves over time. The outputs produced in this research (i.e. the methodology, the granular estimates of skill demand and the preliminary insights) may be useful to both other researchers and policymakers.

In future work, we plan to further refine our methodology to measure skill demand and to produce the current version of Nesta's skills taxonomy. Having access to granular and timely information on skill demand is crucial to provide an actionable evidence base that could be used by policymakers to develop interventions for the benefit of businesses, workers and the economy overall.

References

[1] P. Adrjan and R. Lydon. Clicks and jobs: measuring labour market tightness using online data. *No. 6/EL/19. Central Bank of Ireland.*, 2019.

[2] J. Albrecht, B. Decreuse, and S. Vroman. Directed search with phantom vacancies. 2020.

[3] D. Bosworth. Review of innovative approaches for analysing diverse sources of labour market information. *Internal report*, Forthcoming.

[4] E. Cammeraat and M. Squicciarini. Assessing the properties of burning glass technologies' data to inform use in policy relevant analysis. *Organisation for Economic Co-operation and Development*, 2020.

[5] Centre for Progressive Policy. The data deficit: Why a lack of information undermines the uk skills system, 2018. URL https://www.progressive-policy.net/publications/the-data-deficit-why-a-lack-of-information-undermines-the-uk-skills-system.

[6] A. Cheron and B. Decreuse. Matching with phantoms. *The Review of Economic Studies*, 84(3):1041–1070, 2017.

[7] N. Cominetti, L. Gardiner, and H. Slaughter. The Full Monty. Facing up to the challenge of the coronavirus labour market crisis, 2020. URL https://www.resolutionfoundation.org/publications/the-full-monty/.

[8] S. J. Davis and B. S. de la Parra. Application flows. *Unpublished manuscript*, 2020.

[9] S. J. Davis, R. J. Faberman, and J. C. Haltiwanger. The establishment-level behavior of vacancies and hiring. *The Quarterly Journal of Economics*, 128(2):581–622, 2013.

[10] Department for Education. Employer skills survey 2017: Uk findings, 2018. URL https://www.gov.uk/government/publications/employer-skills-survey-2017-uk-report.

[11] J. Djumalieva and C. Sleeman. An open and data-driven taxonomy of skills extracted from online job adverts. *Developing Skills in a Changing World of Work: Concepts, Measurement and Data Applied in Regional and Local Labour Market Monitoring Across Europe, 425*, 2018.

[12] J. Djumalieva and C. Sleeman. Applications of the data-driven skills taxonomy for measuring skills demand and ons soc revision. *Internal report, submitted to the ONS*, 2019.

[13] J. Djumalieva, S. Garasto, and C. Sleeman. Measuring regional skills mismatches and access to jobs, 2019. URL https://productivityinsightsnetwork.co.uk/app/uploads/2019/08/Nesta_regional_skill_mismatch_reportv2.pdf.

[14] R. J. Faberman and M. Kudlyak. The intensity of job search and search duration. *American Economic Journal: Macroeconomics*, 11(3):327–57, 2019.

[15] L. Gambin, T. Hogarth, L. Murphy, K. Spreadbury, C. Warhurst, and M. Winterbotham. Research to understand the extent, nature and impact of skills mismatches in the economy: May 2016, 2016.

[16] C. Gini. Concentration and dependency ratios. *Rivista di politica economica*, 87:769–792, 1997.

[17] A. Haldane. Is all economics local?, 2019. URL https://www.bankofengland.co.uk/speech/2019/andy-haldane-sheffield-political-economy-research-institute-annual-lecture-2019.

[18] Institute for Employment Studies. Weekly vacancy analysis: Vacancy trends in week-ending 28 june 2020, 2020. URL https://www.employment-studies.co.uk/system/files/resources/files/IES_Briefing_-_we_03072020.pdf.

[19] E. Magrini and P. Adrjan. Where in the UK is it hardest to find a job? *Indeed hiring lab and Centre for cities blog*, 2020. URL https://www.hiringlab.org/uk/blog/2020/07/15/regional-job-competition/.

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[21] J. Nania, H. Bonella, D. Restuccia, and B. Taska. No longer optional: Employer demand for digital skills. *Burning Glass Technologies*, 2019. URL https://www.burning-glass.com/wp-content/uploads/no_longer_optional_report.pdf.

[22] Office for National Statistics. National statistics quality review of employment and jobs statistics, 2006. URL http://webarchive.nationalarchives.gov.uk/20160105160709/http:/www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-reviews/theme/labour-market/nsqr-44/nsqr-series-report-no--44--review-of-employment-and-jobs-statistics-.pdf.

[23] Office for National Statistics. Census geography, 2016. URL https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography#travel-to-work-area.

[24] Office for National Statistics. 2011 census aggregate data, data collection, 2016.

[25] Office for National Statistics. Vacancy survey, data collection., 2019.

[26] Office for National Statistics. Annual population survey, data collection., 2019.

[27] Office for National Statistics. Using adzuna data to derive an indicator of weekly vacancies: Experimental statistics, 2019. URL https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/usingadzunadatatoderiveanindicatorofweeklyvacanciesexperimentalstatistics.

[28] J. Orlik, M. Rhode, R. Douglas, P. Ward, and R. Scott. Finding opportunities in uncertainty, 2020. URL https://media.nesta.org.uk/documents/23-Nesta-Finding_Opportunities_in_Uncertainty.pdf.

[29] N. Reimers and I. Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[30] D. Restuccia, B. Taska, and S. Bittle. Different skills, different gaps: Measuring and closing the skills gap. *Developing Skills in a Changing World of Work: Concepts, Measurement and Data Applied in Regional and Local Labour Market Monitoring Across Europe*, page 207, 2018.

[31] D. Skandalis. Breaking news: information about firms' hiring needs affects the direction of job search. Technical report, mimeo, 2018.

[32] C. Sleeman and J. Djumalieva. Measuring regional skill mismatches: Can big data help?, 2020. URL https://data-viz.nesta.org.uk/measuring-skill-mismatches/index.html.

[33] S. Smit, T. Tacke, S. Lund, J. Manyika, and L. Thiel. The future of work in Europe. 2020.

[34] N. Swier, F. Hajnovic, T. Declite, M. Rengers, C. Islam, I. Jansson, D. Wu, S. Elezovic, C. Grahonja, C. Pierrakou, E. Bisotti, M. Bergat, A. Eidelman, R. Alves, and M. Fernandes. Web scraping / job vacancies, final technical report, 2018.

[35] Tech Nation. UK tech for a changing word, 2020. URL https://technation.io/report2020/.

[36] TheCityUK. Enabling growth across the UK 2019, 2019. URL https://www.thecityuk.com/research/enabling-growth-across-the-uk-2019/.

[37] J. Thurgood, A. Turrell, D. Copple, J. Djumalieva, and B. Speigner. Using online job vacancies to understand the uk labour market from the bottom-up. 2018.

[38] A. Turrell, B. Speigner, J. Djumalieva, D. Copple, and J. Thurgood. Using job vacancies to understand the effects of labour market mismatch on uk output and productivity. 2018.

[39] A. Turrell, B. J. Speigner, J. Djumalieva, D. Copple, and J. Thurgood. Transforming naturally occurring text data into economic statistics: The case of online job vacancy postings. 2019.

Appendix 1. Linking job adverts to ESCO-based skill clusters

To analyse skill demand by category, we first needed to sort the thousands of unique skills mentioned in job adverts into meaningful categories. Previous work by Nesta has generated these categories in a data-driven way by clustering the skills contained in the job adverts (Djumalieva and Sleeman, 2018). In this work, we adopted a different approach to ensure we were still able to capture occupations that are less frequently advertised in online job adverts. We began by taking an expert derived "base" skills taxonomy, developed from ESCO skills which cover all sectors of the economy. We then enriched and broadened this base taxonomy with the additional skills that are mentioned within job adverts[32]. This approach allowed us to blend both data-driven skills (i.e. from job adverts) and expert-derived skills (i.e. from ESCO) into a single taxonomy. The challenge of this approach was assigning the *skills* mentioned in job adverts (which we hereafter refer to as "vacancy skills") into the most appropriate skill clusters in the base taxonomy. This mapping, from skills to skill clusters, is important because we also used it to assign whole *job adverts* to skill clusters. In the following section of this appendix, we describe the first stage in more detail. Our approach to mapping job adverts to skill clusters is instead outlined in the *Methods* section.

Mapping vacancy skills to skill clusters

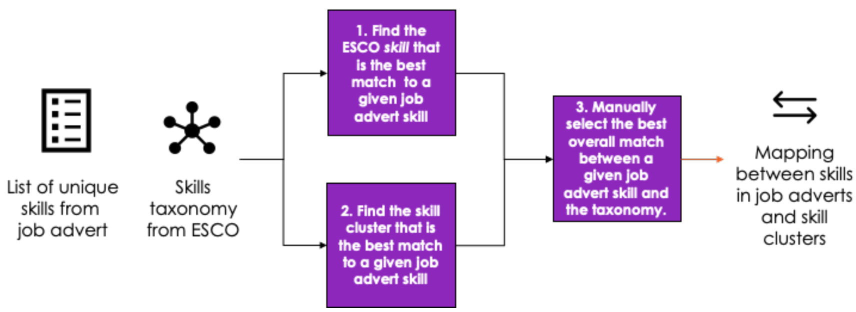


Figure 17: Overview of the process to link skills referenced in job adverts to ESCO-based skills clusters.

Figure A1. Overview of the process to link skills referenced in job adverts to ESCO-based skills clusters.

We used two methods to assign vacancy skills extracted from online job adverts to the base skills taxonomy (Figure A1). First, we looked for the *individual skill* within the ESCO framework that most closely matches each of the vacancy skills. This was done by looking for "almost-exact matches" (e.g. "database management" and "databases management") and semantic similarities[33] *between skills* (e.g. "office supplies management" and "maintain inventory of office supplies"). We then looked directly for the third level *skill cluster* that most accurately matched each of the vacancy skills. This was again based on semantic similarity[34], but *between a skill and a skill cluster* . Finally, we manually reviewed the matches obtained for each of the vacancy skills (around 10,000 skills) to find the best fit overall. We acted conservatively and only kept the matches that we could validate. We were able to match 60% of the vacancy skills, although this percentage increases to 87% among the most frequent 1700 vacancy skills, which together cover 95% of total skill appearances.

[32]Some of these skills will overlap with skills contained within the ESCO framework. However, our algorithm was able to match these skills together. We also did not include skills that are tagged as "soft skills" by TK because, at the moment, there is no single suitable skill cluster.

[33]Semantic similarity was computed as the cosine similarity between the sentence embeddings (Reimers and Gurevych, 2019) of the skills we were comparing. Cosine similarity ranges between 0 (two vectors have no similarities) and 1 (two vectors are exactly the same). The skill from the ESCO framework with the highest similarity with the target vacancy skill was taken as the best match (provided the similarity is higher than 0.7).

[34]Semantic similarity was computed as the cosine similarity between the sentence embeddings (Reimers and Gurevych, 2019) of the target vacancy skills and representative sentence embeddings for the skill clusters. The latter were computed as the (weighted) average sentence embedding across the skills comprising the cluster. The skill cluster within the taxonomy with the highest similarity with the target vacancy skill was taken as the best match (provided the similarity is higher than 0.6).

A limitation of this approach is that it is not always straightforward to match the two groups of skills. For example, some skills might look similar but have different meanings ("image processing" vs "analysis of images"). As a result, not all vacancy skills can be inserted into the taxonomy[35]. This might introduce some bias into the breakdown of skill demand by cluster, since it depends on both the degree to which the skill cluster is represented in online job adverts and the number of vacancy skills that are assigned to that cluster. To quantify the effects more precisely, we could manually assign unmatched skills to skill clusters for the uncategorised skills. In future work, we could also increase the number of successful matches by adding context to the vacancy skills, for example by linking them to Wikipedia entries or by creating new additional skill clusters.

[35] These skills were therefore not included in the rest of the analysis.

Appendix 2. Mapping to SIC 2007

This section provides more details on the approach taken to map job adverts to industries (SIC 2007). First, we describe how we group together some of the sections in the UK industrial classification. Then, we add some technical details regarding the way we built the crosswalk from SOC to SIC codes.

To maximise the number of job adverts that we can link to SIC codes by crosswalking from TK's own industrial categories, we grouped some industries together based on their relationship with these categories Specifically, we created the following groups. "Electricity, gas, steam and air conditioning supply" (D) and "Water supply, sewerage, waste management and remediation activities" (E) were combined into "Utilities (energy, water and waste)". "Professional, scientific and technical activities"(M) and "Education" (P) were grouped into "Educational and professional activities". "Real estate activities" (L), "Public administration and defence; compulsory social security" (O) and "Other service activities" (S) were combined into "Personal and public services".

SOC to SIC crosswalk

This section describes in more detail the steps taken to build a direct match between SOC codes and SIC codes. Specifically, we used a four-fold approach:

1. Word matching using the section SIC code labels.
2. Token similarity using the four digit SIC code descriptions.
3. Lemmatising and matching using the four digit SIC code descriptions.
4. Manually matching a few SOC codes which did not have any direct match to a SIC code description (e.g. draughtsperson).

We started by identifying relevant keywords in the section SIC code labels, such as "agriculture", "manufacturing" or "construction". We then looked for these keywords within SOC code labels: a positive match meant that we could directly assign a SOC code to the SIC code with the corresponding keyword. The next step was to use spacy and their token similarity function to calculate the similarity between the SOC code description and the four digit SIC code descriptions, with stopwords removed. A threshold similarity was set to 0.25, based on a manual check of the data, and any match above this was assigned. In some cases, this led to multiple SIC codes being matched, therefore we combined the token similarity with lemmatisation. By reducing the words to their root form in both the SOC code descriptions and the SIC four-digit descriptions, we identified any exact matches. This also led to multiple SIC letters being assigned. We then took the counts of the SIC letters for the combined token similarity and lemmatised matches and took the most common SIC letter to be the best estimate for the SOC code to SIC code conversion. Finally, a few of the SOC labels did not match any SIC descriptions. For the remaining SOC labels, we manually assigned a SIC letter.

Appendix 3. Average share of demand by occupational groups

In this section, we show the composition of the stock of vacancies by occupational groups (SOC 2010) at multiple levels of granularity, averaged across all years between 2015 and 2019 (percentage).

Sub-major groups

Table A1. Share of vacancies and employment for sub-major occupational groups (SOC 2010). The indicators were averaged across all years between 2015 and 2019, together with their respective share of employment [26].

Sub-major group	Occupation	Vacancies share (%)	Employment share (%)
11	Corporate managers and directors	10.65	7.63
12	Other managers and proprietors	3.09	3.16
21	Science, research, engineering and technology professionals	11.17	5.6
22	Health professionals	5.72	4.43
23	Teaching and educational professionals	4.69	4.94
24	Business, media and public service professionals	9.07	5.54
31	Science, engineering and technology associate professionals	3.83	1.86
32	Health and social care associate professionals	0.31	1.52
33	Protective service occupations	0.2	1.14
34	Culture, media and sports occupations	1.5	2.44
35	Business and public service associate professionals	13.41	7.42
41	Administrative occupations	6.72	8.11
42	Secretarial and related occupations	1.85	2.15
51	Skilled agricultural and related trades	0.19	1.16
52	Skilled metal, electrical and electronic trades	2.32	3.67
53	Skilled construction and building trades	1.17	3.47
54	Textiles, printing and other skilled trades	2.6	2.09
61	Caring personal service occupations	4.43	7.13
62	Leisure, travel and related personal service occupations	1.52	2.03
71	Sales occupations	2.92	5.63
72	Customer service occupations	4.31	1.93
81	Process, plant and machine operatives	0.91	2.64
82	Transport and mobile machine drivers and operatives	2.34	3.74
91	Elementary trades and related occupations	0.93	1.65
92	Elementary administration and service occupations	4.16	8.93

Minor groups

Table A2. Share of vacancies and employment for major occupational groups (SOC 2010). The indicators were averaged across all years between 2015 and 2019, together with their respective share of employment [26].

Minor group	**Occupation**	**Vacancies share (%)**	**Employment share (%)**
111	Chief executives and senior officials	0.37	0.32
112	Production managers and directors	0.45	1.62
113	Functional managers and directors	7.71	3.33
115	Financial institution managers and directors	0.04	0.28
116	Managers and directors in transport and logistics	0.37	0.59
117	Senior officers in protective services	0.12	0.16
118	Health and social services managers and directors	0.34	0.3
119	Managers and directors in retail and wholesale	1.26	1.03
121	Managers and proprietors in agriculture related services	0.06	0.12
122	Managers and proprietors in hospitality and leisure services	0.79	0.93
124	Managers and proprietors in health and care services	0.58	0.26
125	Managers and proprietors in other services	1.67	1.85
211	Natural and social science professionals	0.54	0.69
212	Engineering professionals	3.36	1.5
213	Information technology and telecommunications professionals	6.63	3.07
214	Conservation and environment professionals	0.09	0.18
215	Research and development managers	0.54	0.16
221	Health professionals	1.63	1.76
222	Therapy professionals	0.53	0.52
223	Nursing and midwifery professionals	3.56	2.15
231	Teaching and educational professionals	4.69	4.94
241	Legal professionals	1.56	0.65
242	Business, research and administrative professionals	4.61	2.37
243	Architects, town planners and surveyors	0.76	0.82
244	Welfare professionals	0.71	0.6
245	Librarians and related professionals	0.04	0.13
246	Quality and regulatory professionals	1.02	0.45
247	Media professionals	0.37	0.53
311	Science, engineering and production technicians	2.55	1.01
312	Draughtspersons and related architectural technicians	0.22	0.2
313	Information technology technicians	1.06	0.65

Minor group	Occupation	Vacancies share (%)	Employment share (%)
321	Health associate professionals	0.24	0.53
323	Welfare and housing associate professionals	0.07	0.99
331	Protective service occupations	0.2	1.14
341	Artistic, literary and media occupations	0.5	1.33
342	Design occupations	0.62	0.54
344	Sports and fitness occupations	0.37	0.57
351	Transport associate professionals	0.05	0.14
352	Legal associate professionals	0.52	0.23
353	Business, finance and related associate professionals	3.65	2.36
354	Sales, marketing and related associate professionals	7.07	3.01
355	Conservation and environmental associate professionals	0	0.03
356	Public services and other associate professionals	2.13	1.65
411	Administrative occupations: government and related organisations	0.05	1.1
412	Administrative occupations: finance	1.8	2.44
413	Administrative occupations: records	2.19	1.26
415	Other administrative occupations	1.82	2.63
416	Administrative occupations: office managers and supervisors	0.86	0.68
421	Secretarial and related occupations	1.85	2.14
511	Agricultural and related trades	0.19	1.16
521	Metal forming, welding and related trades	0.24	0.32
522	Metal machining, fitting and instrument making trades	0.26	1.01
523	Vehicle trades	0.83	0.84
524	Electrical and electronic trades	0.71	1.4
525	Skilled metal, electrical and electronic trades supervisors	0.29	0.12
531	Construction and building trades	0.79	2.62
532	Building finishing trades	0.19	0.66
533	Construction and building trades supervisors	0.19	0.19
541	Textiles and garments trades	0.02	0.16
542	Printing trades	0.04	0.15
543	Food preparation and hospitality trades	2.51	1.46
544	Other skilled trades	0.03	0.32
612	Childcare and related personal services	1.65	2.63
613	Animal care and control services	0.13	0.32
614	Caring personal services	2.64	4.17
621	Leisure and travel services	0.4	0.6
622	Hairdressers and related services	0.29	0.84
623	Housekeeping and related services	0.64	0.35

Minor group	Occupation	Vacancies share (%)	Employment share (%)
624	Cleaning and housekeeping managers and supervisors	0.18	0.24
711	Sales assistants and retail cashiers	2.64	4.52
712	Sales related occupations	0.25	0.54
713	Sales supervisors	0.03	0.58
721	Customer service occupations	3.3	1.44
722	Customer service managers and supervisors	1.01	0.48
811	Process operatives	0.09	0.82
812	Plant and machine operatives	0.54	0.48
813	Assemblers and routine operatives	0.11	0.81
814	Construction operatives	0.17	0.53
821	Road transport drivers	1.73	2.97
822	Mobile machine drivers and operatives	0.59	0.5
823	Other drivers and transport operatives	0.02	0.27
911	Elementary agricultural occupations	0.02	0.28
912	Elementary construction occupations	0.14	0.54
913	Elementary process plant occupations	0.77	0.82
921	Elementary administration occupations	0.16	0.61
923	Elementary cleaning occupations	1.15	2.26
924	Elementary security occupations	0.6	0.98
925	Elementary sales occupations	0.01	0.41
926	Elementary storage occupations	0.12	1.38
927	Other elementary services occupations	2.1	3.28

Unit groups

Share of vacancies and employment for major occupational groups (SOC 2010). The indicators were averaged across all years between 2015 and 2019, together with their respective share of employment [26].

Unit group	Occupation	Vacancies share (%)	Employment share (%)
1115	Chief executives and senior officials	0.37	0.28
1116	Elected officers and representatives	0	0.04
1121	Production managers and directors in manufacturing	0.24	0.97
1122	Production managers and directors in construction	0.16	0.61
1123	Production managers and directors in mining and energy	0.04	0.05
1131	Financial managers and directors	0.76	0.99
1132	Marketing and sales directors	3.66	0.74
1134	Advertising and public relations directors	0.39	0.11
1135	Human resource managers and directors	1.03	0.58

Unit group	Occupation	Vacancies share (%)	Employment share (%)
1136	Information technology and telecommunications directors	0.65	0.32
1139	Functional managers and directors n.e.c.	1.21	0.39
1150	Financial institution managers and directors	0.04	0.28
1161	Managers and directors in transport and distribution	0.06	0.26
1162	Managers and directors in storage and warehousing	0.31	0.33
1171	Officers in armed forces	0.07	0.09
1172	Senior police officers	0.02	0.03
1173	Senior officers in fire, ambulance, prison and related services	0.03	0.04
1181	Health services and public health managers and directors	0.3	0.2
1184	Social services managers and directors	0.03	0.1
1190	Managers and directors in retail and wholesale	1.26	1.03
1211	Managers and proprietors in agriculture and horticulture	0.06	0.08
1213	Managers and proprietors in forestry, fishing and related services	0	0.05
1221	Hotel and accommodation managers and proprietors	0.26	0.18
1223	Restaurant and catering establishment managers and proprietors	0.41	0.41
1225	Leisure and sports managers	0.11	0.19
1226	Travel agency managers and proprietors	0	0.03
1241	Health care practice managers	0.07	0.08
1242	Residential, day and domiciliary care managers and proprietors	0.5	0.17
1251	Property, housing and estate managers	0.41	0.56
1252	Garage managers and proprietors	0	0.09
1254	Shopkeepers and proprietors – wholesale and retail	0.03	0.42
1255	Waste disposal and environmental services managers	0.02	0.04
1259	Managers and proprietors in other services n.e.c.	1.2	0.66
2111	Chemical scientists	0.08	0.09
2112	Biological scientists and biochemists	0.1	0.28
2113	Physical scientists	0.05	0.09
2114	Social and humanities scientists	0.03	0.07
2119	Natural and social science professionals n.e.c.	0.28	0.16
2121	Civil engineers	0.5	0.26
2122	Mechanical engineers	0.59	0.24

Unit group	Occupation	Vacancies share (%)	Employment share (%)
2123	Electrical engineers	0.48	0.15
2124	Electronics engineers	0.22	0.1
2126	Design and development engineers	0.05	0.24
2127	Production and process engineers	0.3	0.15
2129	Engineering professionals n.e.c.	1.22	0.37
2133	IT specialist managers	0.45	0.61
2134	IT project and programme managers	0.24	0.24
2135	IT business analysts, architects and systems designers	2.07	0.4
2136	Programmers and software development professionals	1.94	1.02
2137	Web design and development professionals	0.71	0.2
2139	Information technology and telecommunications professionals n.e.c.	1.22	0.6
2141	Conservation professionals	0.01	0.05
2142	Environment professionals	0.09	0.13
2150	Research and development managers	0.54	0.16
2211	Medical practitioners	0.92	0.85
2212	Psychologists	0.13	0.11
2213	Pharmacists	0.13	0.19
2214	Ophthalmic opticians	0.1	0.06
2215	Dental practitioners	0.13	0.12
2216	Veterinarians	0.09	0.07
2217	Medical radiographers	0	0.1
2219	Health professionals n.e.c.	0.13	0.2
2221	Physiotherapists	0.19	0.18
2222	Occupational therapists	0.2	0.13
2223	Speech and language therapists	0.1	0.05
2229	Therapy professionals n.e.c.	0.05	0.16
2231	Nurses	3.54	2.01
2232	Midwives	0.01	0.14
2311	Higher education teaching professionals	0.62	0.54
2312	Further education teaching professionals	0.2	0.35
2314	Secondary education teaching professionals	1.88	1.3
2315	Primary and nursery education teaching professionals	1.37	1.38
2316	Special needs education teaching professionals	0.19	0.25
2317	Senior professionals of educational establishments	0.12	0.32
2318	Education advisers and school inspectors	0.05	0.11
2319	Teaching and other educational professionals n.e.c.	0.27	0.68
2412	Barristers and judges	0.43	0.09
2419	Legal professionals n.e.c.	1.13	0.17

Unit group	Occupation	Vacancies share (%)	Employment share (%)
2421	Chartered and certified accountants	1.39	0.6
2423	Management consultants and business analysts	1.25	0.58
2424	Business and financial project management professionals	1.61	0.73
2425	Actuaries, economists and statisticians	0.13	0.14
2426	Business and related research professionals	0.05	0.15
2429	Business, research and administrative professionals n.e.c.	0.18	0.17
2431	Architects	0.33	0.17
2432	Town planning officers	0.04	0.07
2434	Chartered surveyors	0.18	0.19
2436	Construction project managers and related professionals	0.22	0.24
2442	Social workers	0.66	0.32
2443	Probation officers	0.02	0.04
2444	Clergy	0.02	0.15
2449	Welfare professionals n.e.c.	0.01	0.09
2451	Librarians	0.01	0.08
2452	Archivists and curators	0.02	0.05
2461	Quality control and planning engineers	0.18	0.11
2462	Quality assurance and regulatory professionals	0.7	0.31
2463	Environmental health professionals	0.14	0.03
2471	Journalists, newspaper and periodical editors	0.05	0.24
2472	Public relations professionals	0.19	0.17
2473	Advertising accounts managers and creative directors	0.13	0.12
3111	Laboratory technicians	0.1	0.25
3112	Electrical and electronics technicians	0.06	0.1
3113	Engineering technicians	0.17	0.29
3114	Building and civil engineering technicians	0.02	0.07
3115	Quality assurance technicians	0.06	0.09
3116	Planning, process and production technicians	0.83	0.07
3119	Science, engineering and production technicians n.e.c.	1.32	0.13
3121	Architectural and town planning technicians	0.02	0.08
3122	Draughtspersons	0.19	0.12
3131	IT operations technicians	0.06	0.34
3132	IT user support technicians	1	0.31
3213	Paramedics	0.03	0.09
3216	Dispensing opticians	0.07	0.03
3217	Pharmaceutical technicians	0	0.1
3218	Medical and dental technicians	0.13	0.13

Unit group	Occupation	Vacancies share (%)	Employment share (%)
3219	Health associate professionals n.e.c.	0.01	0.19
3231	Youth and community workers	0	0.22
3234	Housing officers	0.01	0.16
3235	Counsellors	0.06	0.07
3239	Welfare and housing associate professionals n.e.c.	0	0.41
3311	Ncos and other ranks	0.01	0.17
3312	Police officers (sergeant and below)	0.07	0.5
3313	Fire service officers (watch manager and below)	0.01	0.12
3314	Prison service officers (below principal officer)	0.03	0.12
3315	Police community support officers	0	0.05
3319	Protective service associate professionals n.e.c.	0.1	0.18
3411	Artists	0.11	0.18
3412	Authors, writers and translators	0.11	0.27
3413	Actors, entertainers and presenters	0.03	0.15
3414	Dancers and choreographers	0	0.06
3415	Musicians	0.01	0.15
3416	Arts officers, producers and directors	0.15	0.27
3417	Photographers, audio-visual and broadcasting equipment operators	0.08	0.25
3421	Graphic designers	0.13	0.29
3422	Product, clothing and related designers	0.49	0.25
3441	Sports players	0.01	0.05
3442	Sports coaches, instructors and officials	0.37	0.32
3443	Fitness instructors	0	0.2
3511	Air traffic controllers	0.01	0.02
3512	Aircraft pilots and flight engineers	0.02	0.08
3513	Ship and hovercraft officers	0.02	0.05
3520	Legal associate professionals	0.52	0.22
3531	Estimators, valuers and assessors	0.51	0.2
3532	Brokers	0.08	0.16
3533	Insurance underwriters	0.09	0.1
3534	Finance and investment analysts and advisers	1.18	0.64
3535	Taxation experts	0.05	0.1
3536	Importers and exporters	0.07	0.03
3537	Financial and accounting technicians	0.82	0.08
3538	Financial accounts managers	0.67	0.52
3539	Business and related associate professionals n.e.c.	0.18	0.53
3541	Buyers and procurement officers	0.94	0.19
3542	Business sales executives	0.45	0.41
3543	Marketing associate professionals	1.1	0.6
3544	Estate agents and auctioneers	0.38	0.16
3545	Sales accounts and business development managers	3.97	1.43

Unit group	Occupation	Vacancies share (%)	Employment share (%)
3546	Conference and exhibition managers and organisers	0.22	0.22
3550	Conservation and environmental associate professionals	0	0.03
3561	Public services associate professionals	0	0.3
3562	Human resources and industrial relations officers	1.54	0.46
3563	Vocational and industrial trainers and instructors	0.15	0.49
3564	Careers advisers and vocational guidance specialists	0.05	0.09
3565	Inspectors of standards and regulations	0.07	0.13
3567	Health and safety officers	0.32	0.18
4112	National government administrative occupations	0.03	0.51
4113	Local government administrative occupations	0.01	0.45
4114	Officers of non-governmental organisations	0.02	0.14
4121	Credit controllers	0.07	0.12
4122	Book-keepers, payroll managers and wages clerks	0.3	1.33
4123	Bank and post office clerks	0.04	0.33
4124	Finance officers	0.69	0.12
4129	Financial administrative occupations n.e.c.	0.7	0.54
4131	Records clerks and assistants	0.1	0.36
4132	Pensions and insurance clerks and assistants	0.07	0.22
4133	Stock control clerks and assistants	0.9	0.3
4134	Transport and distribution clerks and assistants	0.34	0.2
4135	Library clerks and assistants	0.02	0.09
4138	Human resources administrative occupations	0.76	0.1
4159	Other administrative occupations n.e.c.	1.82	2.39
4161	Office managers	0.86	0.55
4211	Medical secretaries	0.14	0.21
4212	Legal secretaries	0.2	0.12
4213	School secretaries	0.03	0.22
4214	Company secretaries	0.03	0.11
4215	Personal assistants and other secretaries	0.71	0.61
4216	Receptionists	0.65	0.75
4217	Typists and related keyboard occupations	0.09	0.12
5111	Farmers	0.04	0.42
5112	Horticultural trades	0.02	0.06
5113	Gardeners and landscape gardeners	0.01	0.5

Unit group	Occupation	Vacancies share (%)	Employment share (%)
5114	Groundsmen and greenkeepers	0.11	0.09
5119	Agricultural and fishing trades n.e.c.	0.01	0.08
5211	Smiths and forge workers	0	0.02
5212	Moulders, core makers and die casters	0.01	0.01
5213	Sheet metal workers	0.02	0.05
5214	Metal plate workers, and riveters	0	0.02
5215	Welding trades	0.17	0.2
5216	Pipe fitters	0.03	0.03
5221	Metal machining setters and setter-operators	0	0.16
5222	Tool makers, tool fitters and markers-out	0.03	0.04
5223	Metal working production and maintenance fitters	0.15	0.68
5224	Precision instrument makers and repairers	0	0.08
5225	Air-conditioning and refrigeration engineers	0.07	0.05
5231	Vehicle technicians, mechanics and electricians	0.75	0.55
5232	Vehicle body builders and repairers	0.02	0.09
5234	Vehicle paint technicians	0.03	0.04
5235	Aircraft maintenance and related trades	0.01	0.09
5236	Boat and ship builders and repairers	0.01	0.04
5237	Rail and rolling stock builders and repairers	0	0.03
5241	Electricians and electrical fitters	0.6	0.8
5242	Telecommunications engineers	0.02	0.19
5244	Tv, video and audio engineers	0	0.03
5245	IT engineers	0.08	0.12
5249	Electrical and electronic trades n.e.c.	0.02	0.25
5250	Skilled metal, electrical and electronic trades supervisors	0.29	0.11
5311	Steel erectors	0	0.03
5312	Bricklayers and masons	0.07	0.24
5313	Roofers, roof tilers and slaters	0.01	0.15
5314	Plumbers and heating and ventilating engineers	0.32	0.57
5315	Carpenters and joiners	0.21	0.73
5316	Glaziers, window fabricators and fitters	0.01	0.13
5319	Construction and building trades n.e.c.	0.16	0.78
5321	Plasterers	0.03	0.16
5322	Floorers and wall tilers	0.02	0.11
5323	Painters and decorators	0.14	0.39
5330	Construction and building trades supervisors	0.19	0.19
5411	Weavers and knitters	0	0.01
5412	Upholsterers	0.01	0.05

Unit group	Occupation	Vacancies share (%)	Employment share (%)
5413	Footwear and leather working trades	0	0.02
5414	Tailors and dressmakers	0.01	0.04
5419	Textiles, garments and related trades n.e.c.	0	0.04
5421	Pre-press technicians	0.01	0.01
5422	Printers	0.02	0.1
5423	Print finishing and binding workers	0.01	0.04
5431	Butchers	0.02	0.11
5432	Bakers and flour confectioners	0.1	0.1
5433	Fishmongers and poultry dressers	0.01	0.03
5434	Chefs	1.69	0.77
5435	Cooks	0.48	0.24
5436	Catering and bar managers	0.21	0.22
5441	Glass and ceramics makers, decorators and finishers	0	0.04
5442	Furniture makers and other craft woodworkers	0.01	0.11
5443	Florists	0.01	0.04
5449	Other skilled trades n.e.c.	0.01	0.13
6122	Childminders and related occupations	0.62	0.35
6126	Educational support assistants	1.04	0.51
6131	Veterinary nurses	0	0.06
6132	Pest control officers	0.02	0.02
6139	Animal care services occupations n.e.c.	0.11	0.25
6141	Nursing auxiliaries and assistants	2.1	1
6142	Ambulance staff (excluding paramedics)	0	0.07
6143	Dental nurses	0.1	0.16
6145	Care workers and home carers	0.43	2.41
6146	Senior care workers	0	0.26
6148	Undertakers, mortuary and crematorium assistants	0.02	0.08
6211	Sports and leisure assistants	0.09	0.21
6212	Travel agents	0.27	0.11
6214	Air travel assistants	0.02	0.15
6215	Rail travel assistants	0.01	0.05
6219	Leisure and travel service occupations n.e.c.	0.02	0.08
6221	Hairdressers and barbers	0.07	0.53
6222	Beauticians and related occupations	0.22	0.3
6231	Housekeepers and related occupations	0.49	0.13
6232	Caretakers	0.16	0.23
6240	Cleaning and housekeeping managers and supervisors	0.18	0.24
7111	Sales and retail assistants	2.12	3.43
7112	Retail cashiers and check-out operators	0.08	0.6
7113	Telephone salespersons	0.05	0.11
7114	Pharmacy and other dispensing assistants	0.11	0.26

Unit group	Occupation	Vacancies share (%)	Employment share (%)
7115	Vehicle and parts salespersons and advisers	0.28	0.12
7121	Collector salespersons and credit agents	0.01	0.04
7122	Debt, rent and other cash collectors	0.05	0.08
7123	Roundspersons and van salespersons	0.02	0.08
7124	Market and street traders and assistants	0	0.06
7125	Merchandisers and window dressers	0.06	0.08
7129	Sales related occupations n.e.c.	0.11	0.19
7130	Sales supervisors	0.03	0.58
7211	Call and contact centre occupations	0.26	0.33
7213	Telephonists	0.04	0.04
7214	Communication operators	0.01	0.11
7215	Market research interviewers	0.07	0.04
7219	Customer service occupations n.e.c.	2.91	0.92
7220	Customer service managers and supervisors	1.01	0.48
8111	Food, drink and tobacco process operatives	0.01	0.47
8112	Glass and ceramics process operatives	0	0.01
8113	Textile process operatives	0.01	0.03
8114	Chemical and related process operatives	0.02	0.13
8115	Rubber process operatives	0	0.01
8116	Plastics process operatives	0.01	0.07
8117	Metal making and treating process operatives	0	0.03
8118	Electroplaters	0	0.02
8119	Process operatives n.e.c.	0.05	0.04
8121	Paper and wood machine operatives	0	0.08
8122	Coal mine operatives	0.05	0
8123	Quarry workers and related operatives	0.01	0.03
8124	Energy plant operatives	0.01	0.02
8125	Metal working machine operatives	0.39	0.19
8126	Water and sewerage plant operatives	0.03	0.03
8127	Printing machine assistants	0.02	0.03
8129	Plant and machine operatives n.e.c.	0.02	0.09
8131	Assemblers (electrical and electronic products)	0.02	0.09
8132	Assemblers (vehicles and metal goods)	0.02	0.14
8133	Routine inspectors and testers	0.03	0.22
8134	Weighers, graders and sorters	0	0.06
8135	Tyre, exhaust and windscreen fitters	0.03	0.04
8137	Sewing machinists	0.02	0.1
8139	Assemblers and routine operatives n.e.c.	0	0.15
8141	Scaffolders, stagers and riggers	0.02	0.09
8142	Road construction operatives	0.01	0.07

Unit group	Occupation	Vacancies share (%)	Employment share (%)
8143	Rail construction and maintenance operatives	0.01	0.03
8149	Construction operatives n.e.c.	0.13	0.33
8211	Large goods vehicle drivers	1.47	0.96
8212	Van drivers	0	0.81
8213	Bus and coach drivers	0.04	0.38
8214	Taxi and cab drivers and chauffeurs	0.18	0.72
8215	Driving instructors	0.03	0.11
8221	Crane drivers	0.1	0.04
8222	Fork-lift truck drivers	0.33	0.28
8223	Agricultural machinery drivers	0.01	0.03
8229	Mobile machine drivers and operatives n.e.c.	0.15	0.15
8231	Train and tram drivers	0.01	0.09
8232	Marine and waterways transport operatives	0.01	0.02
8233	Air transport operatives	0	0.05
8234	Rail transport operatives	0	0.05
8239	Other drivers and transport operatives n.e.c.	0	0.07
9111	Farm workers	0.01	0.19
9112	Forestry workers	0.01	0.02
9120	Elementary construction occupations	0.14	0.55
9132	Industrial cleaning process occupations	0.04	0.09
9134	Packers, bottlers, canners and fillers	0.06	0.45
9139	Elementary process plant occupations n.e.c.	0.67	0.29
9211	Postal workers, mail sorters, messengers and couriers	0.16	0.5
9219	Elementary administration occupations n.e.c.	0.01	0.11
9231	Window cleaners	0.01	0.09
9232	Street cleaners	0	0.03
9233	Cleaners and domestics	1.04	1.82
9234	Launderers, dry cleaners and pressers	0.02	0.08
9235	Refuse and salvage occupations	0.06	0.12
9236	Vehicle valeters and cleaners	0.01	0.09
9241	Security guards and related occupations	0.57	0.57
9242	Parking and civil enforcement occupations	0.03	0.04
9249	Elementary security occupations n.e.c.	0	0.06
9251	Shelf fillers	0.01	0.3
9259	Elementary sales occupations n.e.c.	0	0.11
9260	Elementary storage occupations	0.12	1.38
9271	Hospital porters	0	0.05
9272	Kitchen and catering assistants	0.56	1.54
9273	Waiters and waitresses	0.9	0.85
9274	Bar staff	0.37	0.62

Unit group	Occupation	Vacancies share (%)	Employment share (%)
9275	Leisure and theme park attendants	0	0.1
9279	Other elementary services occupations n.e.c.	0.26	0.11