OECD Local Economic and Employment Development (LEED) Papers

How well do online job postings match national sources in large English speaking countries?

Benchmarking Lightcast data against statistical sources across regions, sectors and occupations

By Alexandra Tsvetkova, Elettra D'Amico, Alexander Lembcke, Polina Knutsson, Wessel Vermeulen

This paper presents the first international assessment of the Lightcast vacancy data representativeness based on benchmarking against officially reported vacancy data in Australia, Canada, the United Kingdom and the United States. The analysis compares distributions in the Lightcast data versus official data across large (TL2) regions, industrial sectors and occupational categories. The analysis shows differences in representativeness across countries and on the three dimensions considered. In general, regional representativeness is considerably better than both occupational and sectoral representativeness.

OECD

ABOUT THE OECD

The OECD is a multi-disciplinary inter-governmental organisation with member countries which engages in its work an increasing number of non-members from all regions of the world. The Organisation's core mission today is to help governments work together towards a stronger, cleaner, fairer global economy. Through its network of specialised committees and working groups, the OECD provides a setting where governments compare policy experiences, seek answers to common problems, identify good practice, and co-ordinate domestic and international policies. More information available: www.oecd.org.

ABOUT OECD LOCAL ECONOMIC AND EMPLOYMENT DEVELOPMENT (LEED) PAPERS

The OECD Local Economic and Employment Development (LEED) Programme Papers present innovative ideas and practical examples on how to boost local development and job creation. A wide range of topics are addressed, such as employment and skills; entrepreneurship; the social economy and social innovation; culture; and local capacity building, among others. The series highlights in particular policies to support disadvantaged places and people, such as the low skilled, the unemployed, migrants, youth and seniors.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome, and may be sent to the Centre for Entrepreneurship, SMEs, Regions and Cities, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This paper is authorised for publication by Lamia Kamal-Chaoui, Director, Centre for Entrepreneurship, SMEs, Regions and Cities, OECD.

This document, as well as any statistical data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Acknowledgements

Table of contents

FIGURES

TABLES

Executive summary

The richness of online job postings data provides a means to have granular and timely analyses of labour demand across occupations, sectors and regions, and skills. Indeed, the level of detail available often surpasses that of official statistics, and, combined with timeliness, this has led to a considerable increase in the use of web-scraped data for policy-analysis in recent years. However, there is relatively little understanding of how representative web-scraped data of online job postings are relative to official data on all vacancies and job openings.

This paper offers an international assessment of the representativeness of web-scraped vacancy data from one provider (Lightcast) for Australia, Canada, the United Kingdom (UK) and the United States (US). It considers distributions across occupational categories, industrial sectors and regions in comparison with official vacancy data from national surveys on vacancies and job openings, covering the period from Q1 2015 to Q2 2022. A companion paper covers European countries for 2019-22. Lightcast started collecting data for most non-English speaking countries much later and compares such data with survey data and administrative statistics of public employment services.

Lightcast data tracks changes in occupation and skills demand over time, also permitting for regional and industry-based aggregations. The analysis using the data can focus on various analytical units, from firms to countries, and information on the skill requirements for jobs can be informative of the diffusion of technology across regions, occupations and industries. Despite this rich information, the data have their limitations mainly due to the method to collect job postings–the use of online job advertisements may not be used equally across occupations, sectors and regions–while the algorithms used to extract location, industry, occupation and skills from the job postings are generally accurate but may contain errors.

The Lightcast data better reflect distributions of vacancies than absolute levels, thus favouring distribution-based analyses. In all four countries, the number of job postings in Lightcast is substantially below the number of job ads of the official data across most sectors and occupations. However, the share of vacancies across sectors and occupations closely follows those in the official data. These shares are also largely stable over time when aggregating the data by quarter.

In Canada, occupations in the categories Business, Finance and Administration, as well as in Management are overrepresented in Lightcast data. Among the four countries covered in this report, Canada is the only country for which vacancies by occupation data can be compared with official statistics, as similar official statistics do not exist in other countries. The shares of these occupations in the Lightcast data exceeded the shares in the official data by 2-6 percentage points (pp) over the period. In turn, five out of ten categories were underrepresented. Over time, patterns were generally stable for most occupations, with the exception of Art, Culture, Recreation, and Sport, which alternated between being underrepresented and overrepresented.

Similar sectors are consistently overrepresented relative to official statistics across the four countries between 2015-22. In Australia, the Education and Health sector is overrepresented by around 15pp in quarter. In the UK the same sector is overrepresented by 12pp and in the US, the Education Healthcare and Social assistance sector is overrepresented by 7pp. Professional, Scientific, Technical and

Administrative services are underrepresented in Australia (10pp), and the US (10pp). The sectoral over- and underrepresentation relative to the official national statistics vary little over time in each country.

Geographically, regional representativeness was very high, with most regions in Australia, Canada, and the US showing only minor discrepancies of 1pp or less (the UK does not report vacancies by region). However, notable deviations include Western Australia (Australia) and Quebec (Canada), which were underrepresented by around 4pp and 13pp, respectively over the period. Over time, spatial representativeness was relatively stable, but some specific regions displayed increased overrepresentation or underrepresentation during different periods.

In summary, while the Lightcast data provide a reasonable reflection of vacancy trends in the four countries, discrepancies in occupational, industrial, and geographical distributions with official statistics exist. Analysts using Lightcast data can ascertain whether such discrepancies affect their findings. Since the under- and overrepresentations of sectors and other dimensions of the data appears to be quite stable over time, statistical techniques, such as data weighting, could be used to correct for it.

1 Introduction

The impact of the COVID-19 pandemic on labour markets was swift and profound. It was also highly uneven across industries, occupations and places. with the situation evolving rapidly. The need to design timely policy responses urgently called for evidence on the changing situation "on the ground". While the national statistical offices ramped up their efforts in real-time data collection, the use of private data sources, particularly in empirical research, expanded too.

Online job postings data aggregated by several companies were widely used to analyse the labour market conditions. The richness of these data and their usefulness, however, extend well beyond understanding of the unfolding trends in labour demand. The details often not available in any official data make the online job postings data an extremely fertile source for various types of economic policy-relevant analyses. Such details include demand for specific skills and very detailed vacancy distributions across places, industries and occupations. The structure of the databases with online job postings allows for cross-tabulation across many dimensions.

Timely knowledge of the specific hiring needs of employers also opens a wide range of opportunities for policy makers to address mismatches. Specific examples of policy interventions that can be highly impactful include: custom-designed training and reskilling programmes to respond to employer demand and megatrends; digitalisation support to SMEs (possibly in specific industries and/or locations); policies to attract workers in specific occupations (e.g. digital nomads); and other initiatives.

In many countries, online job postings data are the only data source reflecting labour demand for occupations, as official statistical agencies tend to report vacancies by industry and/or region. An analysis of labour demand by occupation, especially at a fine geographical scale, offers many insights that would be hard or impossible to obtain from industry data. For example, a simple classification of occupations into routine and non-routine groups depending on the skill content gave rise to a new stream of research on labour market polarisation (Autor, Levy and Murnane, 2003[1]), which helps explain a range of ongoing processes from inequality to falling entrepreneurship rates. Feser (2003[2]) uses information on occupational knowledge from the O*NET database to construct groups of occupations that share the same knowledge bases, the so-called knowledge-based occupational clusters. Linking occupational titles to skill content, Bacolod and co-authors (Bacolod, Blum and Strange, 2009[3]) show that the wage premium varies across different types of skills (e.g. motor skills vs. social skills) and that large cities tend to be slightly more "skilled" relative to smaller cities.

The relevance of occupational data for economic analysis has increased rapidly, since it captures the ongoing changes in the organisation of economic activity, as well as the economic and social implications of the evolving nature of professions, occupations and task combinations. Based on the analysis of occupational data, Duranton and Puga (2005[4]) conclude that cities have shifted from sectoral specialization to functional (i.e. from a setting where headquarters and plants are located in the same place to a decoupling whereby headquarters and business services locate in larger cities, while plants locate in smaller ones). More recent work shows increasing use in US cities of routine-biased technologies, which require skilled workers, and that the Great Recession has accelerated this process (Hershbein and Kahn, 2018[5]) or the emergence of hybrid occupations and the resultant evolution of the job market (Sigelman et al., 2019[6]).

The versatility and the promise of the online job postings data for academic research and policy analysis raises the question of data reliability. In other words, how well do the online job postings data reflect the economic reality? Several attempts to answer this question have been made (both in terms of representativeness and accuracy).[1] The common approach to verify representativeness is to compare data from private providers to that of the official statistical offices. Section 3 briefly overviews previous efforts in this regard.

This paper assesses representativeness of online job postings data from one provider, Lightcast (formerly EMSI Burning Glass and Burning Glass Technologies before that) through comparisons of distributions[2] with the official vacancy data. The comparison covers the four largest English-speaking countries, (in alphabetical order) Australia, Canada, the United Kingdom (UK) and the United States (US), from Q1 2015 to Q2 2022. The analysis presented here is descriptive. The comparisons are made across industries, occupations and regions within each country.[3] Additionally, the evolution of the differences between the official data sources and the Lightcast data over time is considered together with the sensitivity of these differences to the aggregations (monthly, quarterly or annually depending on data availability). Special attention is paid to ensuring comparability of the Lightcast aggregations to the officially reported vacancies through matching the official data collection approach as much as possible, as described in Section 4. A companion paper covers European countries for 2019-22 (Vermeulen and Gutierrez Amaros, 2024[7]).

---

[1] Representativeness refers to the ability of the data to correctly characterise the population of all vacancies in an economy, for instance across industries, occupations, skills, places, etc. Accuracy reflects the ability of the Lightcast algorithm to assign job announcements into correct categories. For example, occupations can be inferred from the job title and, thus, are more likely to be correct. In some cases, however, discerning correct meaning is not straightforward. Lancaster and co-authors (Lancaster, Mahoney-Nair and Ratcliff, 2019[10]) cite difficulties that the Lightcast algorithm used to have with the abbreviation "PA", which can mean Physician Assistant or Professional Association, among other meanings. The algorithm was updated after this specific issue was highlighted in prior checks of the database. The assignment of an announcement to an industry relies on the information about the hiring firm, which may or may not be sufficiently informative. This means that the industrial breakdown of vacancies may be less accurate.

[2] An axillary analysis reported in Annex B presents the ratio over time of the *number* of vacancies in the Lightcast data to the official data across occupational categories, industrial sectors and TL2 regions. Overall, the analysis confirms that the Lightcast data are better suited to study distributions of vacancies (shares) than the levels of labour demand.

[3] The comparisons are performed using the best matches given the official data availability. Statistical offices in all four countries report vacancy data by industrial sector. For occupations, vacancy data are available only in Canada. Regional vacancy data are reported by three countries, Australia, Canada and the US. Job postings from the Lightcast database are aggregated to match the official tabulations as closely as possible, and the resultant distributions are plotted against each other and presented throughout this report.

2 The Lightcast data

The Lightcast data are a tabulation of all on-line job postings collected by the data company Lightcast. The information available in the dataset includes job title and description, location, time of posting, industry codes, (broadly defined) required skills, and some other attributes. The dataset described in this report covers Australia, Canada, the UK and the US over Q1 2012-Q2 2022, although the comparative analysis focuses on Q1 2015-Q2 2022.

The Lightcast algorithm

Lightcast uses a web crawling algorithm to collect information on vacancies from online sources in real time, and aims to include the universe of online job postings. The algorithm parses each job announcement into text categories and applies semantic analysis to standardise text entries. The resulting dataset contains around 70 entries for each announcement, including job title, date of posting, location, identifiers of occupation and industry, job requirements, employer's name, salary and job type.

Location variables include region, city and geographical coordinates (if available) and are based on the finest locational detail specified in the job ad. For example, if an ad specifies a city, the name of the city is used to fill data fields for geographical coordinates, city and region. If the most granular location specified is a district within a city, the geographical coordinates are then based on the district.

Industrial identifiers follow the official classification code used in a country. Data are provided at various levels of detail depending on availability. Data on occupations come in two forms. Lightcast provides job titles from the original announcements and an array of standardized job titles. The standardized variables include national occupational codes (at various levels of detail) and Lightcast's own occupational groupings.

Job requirements include the level of education and experience, skills, certificates, licenses and other characteristics listed in the vacancy announcement. Skills in the Lightcast data are defined broadly and can include abilities, knowledge and competencies mentioned in the ad, for example, "Problem Solving", "Java", "Energetic", "Bilingual", "Computer Literacy", "Tax Law".

An important feature of the algorithm is de-duplication of the job announcements. It strives to ensure that the same announcement published on several websites is included only once in the dataset. There is evidence, however, that the problem of duplication may still be present in the Lightcast data (Morgan et al., 2017[8]).[4]

---

[4] A way to mitigate this problem in empirical analysis may be dropping variable(s) that are likely to differ for duplicate vacancies (e.g. the source of the announcement or country for the European data) and removing observations that are identical after the data source variables have been deleted.

Advantages of the data

The advantages of the Lightcast data are their detailed content, granularity, timeliness and the coverage. The data can be successfully used to study how the demand for occupations and skills changes over time and within researcher-defined units of analysis, from firms to countries. The data can also be used to study trends in specific industries in terms of expected expansions and contractions, or an evolution of occupational and skills composition of industries (in specific places). The ability to detect and monitor these processes at a detailed geographical level makes online job postings data a unique source for place-based policy insights.

In terms of content, the data capture actual job vacancies describing labour market trends across industries and occupations. The detail of the collected information potentially allows moving past (often rigid) official classifications to glean insights into more specific developments, such as changes in the use of certain technologies or knowledge domains.

Central to this is information on the required skills contained in the Lightcast data, which captures what employers are actually looking for. Because job requirements listed in the job postings are provided by the firms seeking to hire, the required skills may or may not correspond to those officially listed in occupational manuals. The Lightcast data, therefore, can be indispensable for the study of occupational and industrial evolutions, including for the detection and understanding of hybrid occupations and industries.

The high granularity of the data (at the vacancy level) allows many types of aggregations, from firm-level to larger geographical regions including by industry, occupation, required worker characteristic and skills. The flexibility of choosing not only the spatial scale but also the ways of defining regions opens doors for meaningful international comparisons, which crucially depend on the consistency of regional definitions. This is important, because countries typically have their own function-based regional delineations, which are not necessarily comparable. For instance, Local Authority Districts (LAD) in the UK correspond to the administrative units, while Metropolitan Statistical Areas (MSA) in the US encompass areas with close economic ties.

The timeliness of the Lightcast data is important for the up-to-date analyses and for a more detailed tracking of economic activity at shorter than annual intervals. The official indicators, in some cases offering comparable information to the Lightcast dataset, are often released with a multi-year delay. The timeliness of the data proved particularly important during the COVID-19 pandemic, which triggered massive and unexpected changes in the labour markets with an urgent need for up-to-date information that was not available from conventional sources at the beginning of the pandemic.

Finally, the (purported) exhaustive nature of the Lightcast data is crucial for detecting the trends in all segments of the economy where new jobs are advertised on-line. As the prevalence of the Internet for vacancy announcements is growing, not least so during and after the pandemic, the usefulness of these data for economic and social analyses is increasing.

Limitations of the data

The limitations of the Lightcast data for empirical analyses are rooted in two domains: how employers use the Internet for job announcements and how the Lightcast algorithm processes the ads. The non-standardised nature of the collected information may also complicate analytical work, as meaningful groupings of the data may be a challenging and time-consuming task in certain circumstances.

Not all job openings are included in the Lightcast data. By definition, online job postings are only vacancies advertised online. As a result, the ability of the Lightcast data to reflect the reality precisely is

likely to differ by location, industry and occupation. This concern, however, is likely of decreasing importance as the pandemic should have made the use of Internet for recruiting more likely.

Furthermore, online job announcements might not map exactly to the firm labour demand. For instance, a single job announcement can be used to attract several workers. An ad may indicate higher requirements than a firm demands to limit the number of applicants or a newly posted opening can reflect firm's inability to fill previously advertised vacancies.[5]

Another limitation is that the accuracy of the processed text entries varies across data fields. After comparing the text of a full ad from a stratified random sample of vacancies to the parsed information contained in the Lightcast data for the US, Carnevale, Jayasunder and Repnikov (2014[9]) conclude that the parser is accurate; however, the level of accuracy differs across data fields. In particular, the level of accuracy was greater than 80% for state, city, occupational title and major occupation group. It was about 73% accurate for the six-digit occupational codes. A higher level of accuracy is reported by Hershbein and Kahn (2018[5]), which likely reflects the fact that Lightcast regularly revises the algorithm and applies the updated version retroactively to the previous series.

Finally, job requirements are used to define skills recorded in the Lightcast data. The extraordinary heterogeneity of the requirements that individual job announcements contain complicates categorisations usually needed for analysis. On the one hand, the listed skills in fact include also competencies and other indicators of workers' suitability for a job and often differ from the definitions of skills common in (for example, labour economics) research. On the other hand, many skills essential for a job can be implied but not mentioned in a web posting, for example cooking skills for restaurant chefs are not likely to make it to the job announcement text. Additional semantic analyses could be needed before the Lightcast data can be used to answer some research questions.

Basic data tabulations and checks

Overall, the Lightcast vacancy aggregations are sizable data sets in all English-speaking countries, particularly in the UK and the US. Table 2.1 shows the number of observations in the data sets available to the OECD in August 2022 by country and year.

Table 2.1. Number of observations per country

Year	Australia	Canada	UK	US
2012	633 956	487 251	5 700 177	14 260 844
2013	750 047	733 790	6 744 285	18 300 288
2014	1 093 630	1 315 779	6 090 821	19 240 620
2015	946 461	1 175 424	7 772 632	21 032 278
2016	1 008 080	1 131 653	8 639 924	23 710 576
2017	1 077 600	1 248 337	9 372 975	22 470 978
2018	1 063 466	1 219 683	8 697 955	29 102 073
2019	1 124 321	1 463 804	6 945 605	35 487 110
2020	928 607	1 302 789	6 426 534	36 470 652
2021	1 429 899	1 931 621	9 951 951	45 517 309
Q1-Q2 2022	771 861	1 281 293	6 583 278	28 472 047

Source: OECD calculations based on the Lightcast data.

---

[5] If an ad is posted by a firm again after 60 days, the Lightcast algorithm considers job as new, as the 60-day window is motivated by industry statistics (Carnevale, Jayasundera and Repnikov, 2014[9]).

While the Lightcast data contain a broad range of detailed information, the nature of the data source (web job postings with content and the level of detail determined by the vacancy announcer) together with the nature of the algorithm leave many missing values. The prevalence of the missing data varies across variables, countries and years. Table 2.2 shows the share of missing values for selected variables in the first semester of 2022.

Table 2.2. Share of missing values (%), selected variables in January-June 2022

	AUS	CAN	GBR	USA
Job Date	0	0	0	0
State or Region (UK)	0	0	52	0
Latitude	NA	0	61	3
Longitude	NA	0	61	3
City	3	0	58	3
National occupational code	17	5	0	4
Lightcast occupational group	17	9	9	7
Employer's name	32	9	42	19
Salary range	70	58	53	62
Industry code - two digits	46	32	47	40
Industry code - three digits	49	45	60	40
Industry code - four digits	53	51	62	44
Industry code - five digits	NA	77	94	70
Industry code - six digits	NA	77	NA	71

Source: OECD calculations based on the Lightcast data.

Data checks are recommended before the data are used. The nature and the extent of such checks should be determined by the specific project that calls for the use of the Lightcast data. It appears that announcements with job titles containing words, which match several occupation groups, are more likely to be misclassified by the Lightcast algorithm. For instance, the algorithm is able to understand that a vacancy *Mainframe Infrastructure Architect* is an opening for an ICT worker, while *Architectural Technician* calls for architectural technologists. In some cases, nevertheless, *Sitecore Architect* can be classified within the Business and Public Service Professionals sub-major group of occupations instead of the correct one, Science and Technology Professionals (the example uses the UK Standard Occupation Classification system or SOC). In the analyses that utilise a spatial dimension, cross-checking geographical information, such as city and coordinates, could be of relevance. In some cases, advertisers list a large nearby city as the place of work, while indicating the actual location in the suburbs elsewhere in the job announcement. While the share of the geographical miss-assignment of this kind is likely to be low and happen predominantly in proximity to large agglomerations, in some applications proper corrections can be essential for the validity of results.[6]

---

[6] Lightcast also provides a possibility to browse through the original postings, which might be useful for cross-checks and validation in some (potentially limited) circumstances. The postings have no job identifier, which means they cannot be linked directly to the Lightcast data. It can be possible, however, to narrow down the set of original postings by using (a variety of) filters, but it is not always obvious which posting corresponds to the entry of interest since even very detailed filtering can return more than one original postings.

3 Previous checks of the Lightcast data

Researchers and policy analysts in the past have analysed the accuracy and representativeness of the Lightcast data. The checks normally cross-validate specific entries in the data against other sources (e.g. databases of job openings collected by other vendors, governments, etc.) or vacancy distributions along various dimensions against distributions in the official aggregated data.

Most of the attempts to evaluate the quality of the Lightcast data appear to focus on the US dataset, which is by far the largest of all national databases offered by Lightcast. Lancaster, Mahoney-Nair and Ratcliff (2019[10]) offer a good summary of the findings published up to 2019. The summarised analyses focus on the Lightcast data between 2010 and 2016. Overall, they highlight a number of considerations that a researcher needs to keep in mind. For example, some data fields (employer, educational requirement, experience requirement) had between 40% and 55% of missing values, while the de-duplication procedure used at that time could potentially overlook a substantial share (7.5% in the specific example given in Morgan et al. (2017[8])) of duplicates. In terms of representativeness, the existing analyses list different industries and sectors that are over- or underrepresented, depending on the benchmarking approach. The most frequently cited as overrepresented are healthcare, finance and insurance, education and manufacturing, the most underrepresented are construction and government (Hershbein and Kahn, 2018[5]; Carnevale, Jayasundera and Repnikov, 2014[9]; Mohnen, Berkes and Taska, 2018[11]). A benchmarking exercise for occupations offers a partially consistent picture. The Lightcast data overrepresent business and financial, computer and mathematical, and healthcare occupations. Education, construction, and production occupations are underrepresented (Mohnen, Berkes and Taska, 2018[11]).

So far, there have been very few analyses assessing representativeness of the Lightcast data internationally. A recent study by Cammeraat and Squicciarini (2021[12]) offers a detailed assessment of the statistical properties and distributional characteristics of the data at the occupational level in six English speaking countries – Australia, Canada, New Zealand, Singapore, the UK and the US for the period 2010-2019. The authors compare the distributions across occupations in the Lightcast data to the distributions in the official employment data at the 1- and 2-digit level of the occupational classifications (both ISCO-08 and the national classifications are used). The authors show that the representativeness of the vacancy data differs across occupational groups (generally management, computer and mathematical occupations are overrepresented, while farming, fishing and forestry; construction and extraction are underrepresented) and over time. It appears that the data quality is better for some years compared to other years. Overall, data quality (representativeness) over time appears to be good (or improving) in all countries except for the United States, where data quality is uneven with the last two years of analysis (2018 and 2019) showing relatively high deviation levels.

4 **Methodology**

General overview and data availability

This report assesses representativeness of the Lightcast data by comparing distributions across industries, occupations and regions against the official data on *vacancies*. This is the closest one can get to "comparing apples to apples" when evaluating web-scraped data. Comparing vacancies to vacancies is essential for understanding data quality for more detailed analyses, for instance those involving occupational or industrial detail. As turn-over rates vary, comparisons between official employment and web-scraped vacancy data might prove misleading for some occupations and industries. For example, in Canada correlations between officially reported employment and vacancy *counts* is 0.3 for Natural Resources, Agriculture and Related and 0.2 for Art, Culture, Recreation and Sport. In two occupational groups (Sales and Service; Manufacturing and Utilities) correlations are negative, indicating that more new jobs open when employment contracts.[7] In terms of *shares*, which should be preferred to counts when working with the Lightcast data, the correlation coefficient between employment and vacancies equals or exceeds 0.5 only for three occupational groups (Health; Education, Law and Social, Community and Government Services; Sales and Service). Correlations are much lower (or negative) for all other occupational groups (Annex A).[8]

The comparisons between the Lightcast and the official data are reported in Chapter 5 in two types of plots. The first type shows shares of vacancies by individual groups within the focal comparison dimension (occupations, sectors or regions) as reflected in the two data sources. Such exposition allows benchmarking the differences between the Lightcast and the official data against the size (prevalence) of each individual group.[9] The second type shows the differences between vacancy shares in the Lightcast and the official data over time using quarterly intervals.[10] Additionally, the means and variances of the over-time differences by individual groups within the focal comparison dimension are reported in a tabular format.

---

[7] Pairwise correlation coefficients exceed 0.7 for four occupational groups out of nine (Management; Business, Finance and Administration; Natural and Applied Sciences and Related as well as Education, Law and Social, Community and Government Services).

[8] When looking at industrial sectors (as opposed to occupational groups), correlations between officially reported employment and vacancies data reveal a comparable picture. Correlations for both counts and shares vary considerably across industrial sectors, although correlation coefficients for the national totals (without industrial breakdown) tend to be high indicating that both employment and vacancies covary over time (results not reported for brevity).

[9] For example, the differences between shares of occupations reported by the Lightcast and official data in Canada are about 5 percentage points for both Sales & Service and Trades, Transport & Equipment Operators & Related. Yet, the former accounts for about 35% of vacances in the official data, while the latter for about 15% (see Figure 5.1).

[10] Differences over time using monthly aggregations are shown in Annex F. Annual aggregations are not shown for brevity.

If official vacancy data are not provided by the national statistical offices, no comparisons are made. Cammeraat and Squicciarini (2021[12]) perform a detailed benchmarking exercise based on the *employment* data from the official data sources. Table 4.1 provides an overview of vacancy data availability from the official public sources.

Table 4.1. Public availability of vacancy data from national statistical offices

Country	Industry detail	Occupation detail	Subnational detail	Cross-tabulations
Australia	Yes	No	Yes	No
Canada	Yes	Yes	Yes	Industry-Province/Territory Occupation – Province/Territory
UK	Yes	No	No	No
US	Yes	No	Yes	No

Source: OECD elaboration.

Specifics of vacancy data by country

The methodologies for collecting and reporting data on vacancies differ across countries. For example, the Australian Bureau of Statistics (ABS) reports vacancies on a quarterly basis with the third Friday of the middle month of the quarter as the reference date. Vacancy data in Canada are available at monthly, quarterly and annual intervals with various levels of detail. In the US, the vacancy data are reported monthly and reflect all job openings on the last day of the reference month. The specifics of vacancy definitions and the data collection approach are briefly described below together with indications of the Lightcast data aggregation methodology for subsequent comparisons.

Australia

Australian Bureau of Statistics (ABS) collects information on vacancies through the Job Vacancies Survey (JVS).[11] A job vacancy is defined as a job available for immediate filling on the survey reference date and for which recruitment action has been taken. Recruitment action includes efforts to fill vacancies by advertising, by onsite or online notices, by notifying employment agencies or trade unions and by contacting, interviewing or selecting applicants already registered with the enterprise or organisation. The JVS reference date is the third Friday of the middle month of the quarter, i.e. February, May, August and November. The data relate to the number of existing vacancies (stock) on the actual survey reference date, not monthly or quarterly. The vacancy data in Australia are available by division of the Australian and New Zealand Standard Industrial Classification 2006 code (ANZSIC 2006) excluding Agriculture, Forestry and Fishing. At the subnational level, vacancies are broken down by states and territories. For the comparisons presented in this report, the total number of vacancies posted within a month before the third Friday of the middle month of each quarter (inclusive) during the Q1 2015 – Q2 2022 period was calculated.[12] These counts were used to calculate shares (reported in Chapter 5 for May 2022), differences in shares (reported in Chapter 5 over time) and all other reported variables both at quarterly and annual frequency.

---

[11] URL: Survey of Job Vacancies | Australian Bureau of Statistics (abs.gov.au)

[12] Vacancies posted one month prior to the reference date are considered open for the purposes of this report. It is impossible to infer from the web scraped vacancy data which vacancies are active in any given moment in time. Assuming that the Lightcast counts a vacancy that stays open past the 60-day window as a new one (Carnevale, Jayasundera and Repnikov, 2014[9]), the one month used for aggregations is a reasonable estimate of active vacancies.

Canada

Statistics Canada collects data on vacancies through compulsory Job Vacancy and Wage Survey (JVWS). Job vacancy for the purposes of this survey is any type of job (full-time, part-time, permanent, temporary, casual, or seasonal) that meets all of the following conditions: i) it is vacant on the reference date (first day of the month) or will become vacant during the month; ii) there are tasks to be carried out during the month for the job in question and iii) the employer is actively seeking a worker outside the organization to fill the job. Jobs that need to be filled by workers who are not considered employees (e.g. subcontractors or external consultants) are excluded from vacancy count.

The survey collects information on vacancies by occupation, industry and province, territory and economic region. The available detail of the Canadian vacancy data tends to be higher compared to other English-speaking countries. For example, industry breakdown is available at the 3-digit North-American Industry Classification System (NAICS)[13] and occupation breakdown up to 4-digit National Occupational Classification (NOC)[14], although data quality tends to decrease (or estimates cannot be reported) as the level of detail goes up. Likewise, some data may be available at more detailed geographical level, for example, metropolitan areas or their components, or by other attributes, such as firm size. The frequency of the data can be annual, quarterly and monthly, depending on the series. This report uses the official quarterly data on vacancies reported for the period from Q2 2015 to Q1 2022 (data for Q2 and Q3 in 2020 are missing due to the COVID-19 pandemic). Calculations based on these series are presented in Chapter 5. Additional monthly aggregations as per data availability are reported in Annex F.

United Kingdom

The UK Office for National Statistics (ONS) provides vacancy estimates obtained from the Vacancy Survey (VS)[15] of employers. Vacancies are defined as positions for which employers are actively seeking recruits from outside their business or organisation. The estimates reflect the stock of vacancies across the economy, excluding agriculture, forestry and fishing. The VS is sent to a sample of approximately 6 000 businesses every month. Data on job postings in the UK are available only by the Standard Industrial Classification (SIC) codes (no occupational or regional breakdowns). For this report, the officially reported quarterly counts of vacancies were used for comparisons with the Lightcast data at the quarterly and annual frequencies during the Q1 2015 – Q2 2022 period.

United States

The US Bureau of Labor Statistics provides data on job openings collected through the Job Openings and Labor Turnover Survey (JOLTS).[16] A job is open (vacant) if it satisfies all three conditions: i) a specific position (full-time, part-time, permanent, short-term or seasonal) exists and there is work available for that position; ii) the job could start within 30 days, whether or not the employer can find a suitable candidate during that time and iii) the employer is actively recruiting workers from outside the establishment to fill the position. Active recruiting may imply advertising in newspapers, on television, or on the radio; posting Internet notices, posting "help wanted" signs, networking or making "word-of-mouth" announcements; accepting applications; interviewing candidates; contacting employment agencies; or

---

[13] Statistics Canada. Table 14-10-0326-01 Job vacancies, payroll employees, job vacancy rate, and average offered hourly wage by industry sector, quarterly, unadjusted for seasonality; DOI**:** https://doi.org/10.25318/1410032601-eng.

[14] Statistics Canada. Table 14-10-0328-01 Job vacancies, proportion of job vacancies and average offered hourly wage by selected characteristics, quarterly, unadjusted for seasonality DOI: https://doi.org/10.25318/1410032801-eng.

[15] URL: Vacancy Survey - Office for National Statistics (ons.gov.uk)

[16] URL: Job Openings and Labor Turnover Survey Home Page (bls.gov)

soliciting employees at job fairs, state or local employment offices, or similar sources. The official vacancy data are available by industrial sector (2-digit NAICS codes) and by state for the total non-farm economy. No breakdown by occupation is provided. The JOLTS data on job openings are collected monthly and include all positions that are open on the last business day of the reference month. Aggregations from the Lightcast database are also done monthly.[17] For this report, these monthly data are used for quarterly aggregations presented in Chapter 5. Additional monthly aggregations are reported in Annex F as per data availability. The period under consideration is January 2015 – June 2022.

Additional considerations

Generally, the Lightcast data capture distributions of vacancies better than the absolute levels, therefore, analyses using distributions are preferred. If a researcher chooses to work with the levels, several important caveats need to be kept in mind. First of all, there is sizable underreporting (and in some cases overreporting) for specific occupations, industries and regions (Annex B). For example, in Canada the Lightcast data reflect only about 40% of all vacancies in Health occupations, about 30% of all vacancies in Manufacturing and Utilities occupations and about 20% of all vacancies in Natural Resources, Agriculture and Related Production occupations on average. At the other extreme, Lightcast overrepresents vacancies relative to official sources for Management and Natural by 129% and Applied Sciences and Related occupations 113%.

When looking at sectors, the underrepresentation becomes more striking in some countries. In Australia, the Lightcast data capture only about 7% of all vacancies in Wholesale Trade; 6% in Construction, and 2% in Administrative and Support Services. The most well-represented industries are Public Administration and Safety together with Education and Training. The Lightcast data reflect on average 75% of vacancies in both. In Canada, the data on average reflect 24% of vacancies in Construction, 12% in Agriculture, Forestry, Fishing and Hunting and 11% in both Wholesale Trade and Management of Companies and Enterprises. On the other extreme, the Lightcast data report 104% of vacancies in Finance and Insurance and 85% in Utilities on average. In the UK, the situation is better, but even there less than 45% of vacancies are captured in Manufacturing, Education and Mining and Quarrying while in seven industries out of 18 more than 100% of vacancies are reported on average.[18] In the US, the Lightcast data capture more than 50% of vacancies on average in two industries out of reported 15 (Educational Services at 94% and Finance and Insurance at 63%). In two industries (Government and Wholesale Trade), 10% or less of vacancies are reported.

The Lightcast data also capture relatively few vacancies by region. In Australia, more than 50% of vacancies are reported only for the Australian Capital Territory on average. The situation is better in Canada, where 60% or more of all officially reported vacancies are reflected in the Lightcast data. The range of reported vacancies ranges between 60% and 93% for all provinces except for Saskatchewan, where the rate is 150% on average. In the US, the share of reported vacancies ranges from 20% in West Virginia to 66% in District of Columbia.

Most importantly, high variation over time in the shares of official vacancies reported in the Lightcast data (Annex B) makes it challenging to adjust the data counts to closer correspond to the official statistics. For example, in Canada the Lightcast data report between 71% and 188% of official vacancies in Natural and Applied Sciences occupations depending on the quarter. Overall, in half of all

---

[17] For comparisons presented in this report, we use the same approach as for Australia, where vacancies posted on the target date plus one month prior are assumed to be active.

[18] Arts, Entertainment & Recreation (145%), Wholesale and Retail Trade; Repair of Motor Vehicles and Motor Cycles (137%), Electricity, Gas, Steam and Air Conditioning (135%), Accommodation and Food Service (116%), Information and Communication (111%), Administrative and Support Service as well as Construction (both 104%).

occupations, the Lightcast data can considerably underreport vacancies in one quarter and overreport in another. The variability over time is even starker in some industries in some countries. Reporting of vacancies in Public Administration and Safety in Australia ranges from 31% to 223%; reporting of vacancies in Public Administration in Canada ranges from 38% to 214%. Nevertheless, when it comes to industries, the Lightcast data are more likely to underreport official vacancy counts except in the UK. In Australia, vacancies in 14 out of 18 industries are underreported in every period between Q1 2015 and Q2 2022. In Canada, this is the case for 14 out of 20 industries and in the US, for 13 out of 15 industries. Variability over time is also considerable for regions. The Lightcast data both underreports and overreports (depending on the period) vacancies in seven out of eight Australian regions and in 10 out of 13 regions in Canada. In the US, reported vacancies in the Lightcast data exceed the official statistics in some periods only for the District of Columbia.

5 Representativeness of the Lightcast data

By occupation

This subsection compares distributions in the official vacancy data by occupational category (OC) to that in the Lightcast data. Since the analysis is limited to matching vacancies to vacancies, the comparison can be performed only for Canada due to data (un)availability. The comparisons are shown first for 2022, followed by comparisons over time using quarterly aggregations.

In 2022

Figure 5.1 confirms the previously reported pattern of overrepresentation of management and professional occupations and underrepresentation of the lower skilled ones in Canada. Four occupational categories (Business, Finance and Administration; Management; Natural and Applied Sciences and Related and Education, Law & Social, Community & Government Services) are overrepresented with the differences in the shares ranging from around 2 to 6 percentage points. Five out of ten occupational categories are underrepresented with the differences in shares ranging from 1 to 5 percentage points.

Figure 5.1. Comparison of distributions across occupational categories in 2022, Canada

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
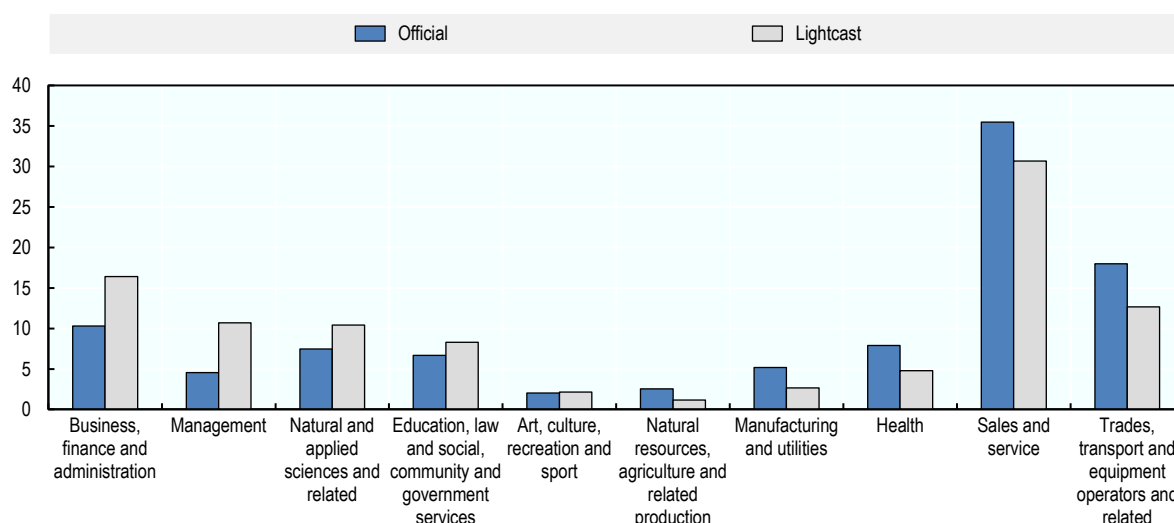
Note: Data for Q2 2022; occupational categories are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey, 2022) data.

Over time

Figure 5.2 reveals relatively stable patterns over time for some occupations and relatively more varying patterns for others. Those stable patterns are observed for Art, Culture, Recreation and Sport; Education, Law and Social, Community and Government Service; Manufacturing and Utilities; and Natural Resources, Agriculture and Related Production. All occupations, except for Art, Culture, Recreation and Sport, are either consistently underrepresented or overrepresented. The magnitude of differences between the shares reported by the Lightcast and official data does not exceed 7 percentage points on average.[19]

Figure 5.2. Representativeness by occupational categories over time, Canada

Differences between vacancy shares reported by the Lightcast data and the official data.

Notes: Data for Q2 and Q3 in 2020 are missing due to the COVID-19 pandemic.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

---

[19] Annual aggregations instead of quarterly ones (not reported for brevity) do not considerably improve the data performance but make more visible changes in data representativeness for some occupations. Between 2015 and 2022, overrepresentation of Natural and Applied Sciences and Related went down from slightly over 6 to under 4 percentage point difference. Underrepresentation decreased for Sales and Service (from -10pp to -6 pp) and Natural Resources, Agriculture and Related Production (from -3pp to -1pp) but increased for Health (from -1pp to -3pp).

Table 5.1. Summary statistics for differences in shares of occupational categories, Canada

Occupational group	Mean	Std. dev.	Min	Max
Management	5.41	1.12	2.41	7.63
Business, Finance & Administration	5.94	1.51	2.29	9.86
Natural & Applied Sciences & Related	5.38	1.36	2.98	9.26
Health	-2.70	1.27	-5.71	-0.80
Education, Law & Social, Community & Government Services	2.19	0.61	1.02	3.95
Art, Culture, Recreation & Sport	0.05	0.23	-0.43	0.38
Sales & Service	-6.45	2.10	-9.64	-2.31
Trades, Transport & Equipment Operators & Related	-4.59	1.75	-8.01	-1.56
Natural Resources, Agriculture & Related Production	-2.56	0.65	-3.87	-1.36
Manufacturing & Utilities	-2.68	0.51	-3.53	-1.76

Note: The table shows summary statistics for each of the series shown in Figure 5.2.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

By industrial sector

This subsection focuses on representativeness by industrial sector. First, comparisons for 2022 are shown followed by comparisons over time using quarterly aggregations (Annex F shows aggregations using monthly intervals as per official data availability). Official vacancy data are reported for all four countries, although there are differences in how sectors are aggregated. The analysis below shows comparisons of distributions across ten sectors[20] for ease of display.[21]

---

[20] The extreme differences shown in the comparisons in many cases should be interpreted as the "upper bound" – differences for disaggregated sectors tend to be lower. This is so because aggregations generally follow proximity of the production processes across industries and those are linked to the likelihood of over- or underrepresentation in the web-scraped data.

[21] Annex C shows how ten sectors were aggregated from the official data. Annex D shows the 2022 comparisons without aggregations and Annex E comparisons over time and without aggregations using quarterly intervals.

Results for 2022

Figure 5.3. Comparison of distributions across sectors in 2022, Australia

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
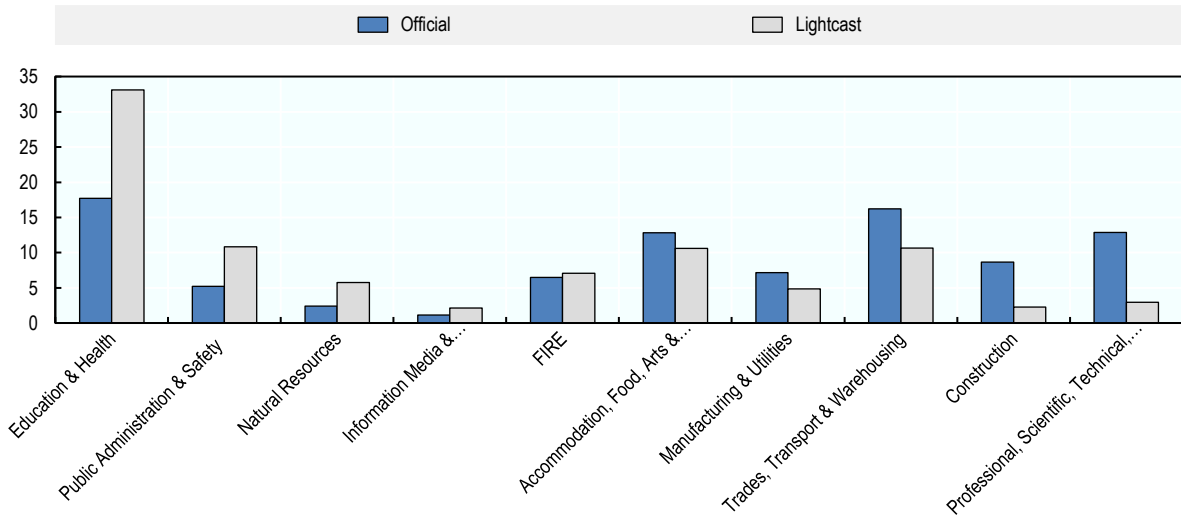
Note: Data for 21 April – 20 May 2022 in the Lightcast data, Q2 2022 in the official data; sectors are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented; FIRE stands for Finance, Insurance and Real Estate.
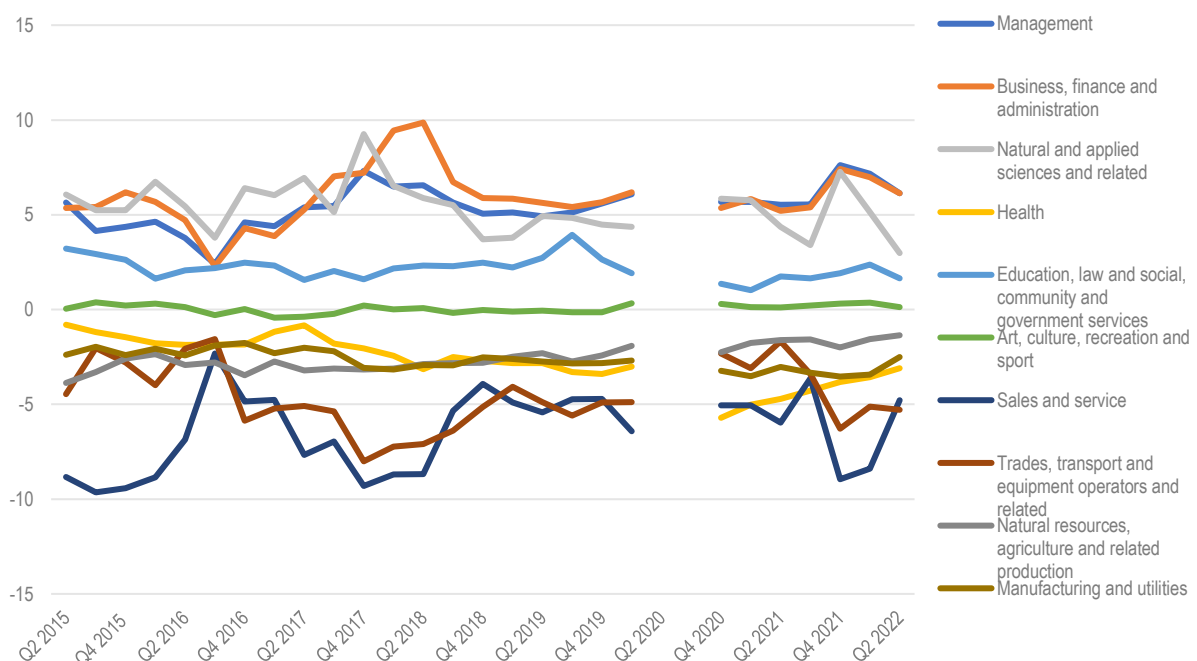Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly, 2022) data.

The differences in shares between the Lightcast and the official data in Australia are the largest among the four countries considered. The share of Lighgtcast vacancies in Education and Health is 15pp higher and the share of vacancies in Professional, Scientific, Technical, Administrative and Other Services is 10pp lower compared to what the Australian Bureau of Statistics reports. Differences of more than five percentage points are observed for Public Administration and Safety (overrepresented); Trade, Transport and Warehousing as well as Construction (both underrepresented) (Figure 5.3). The difference of three percentage points for Natural Resources (overrepresentation in the Lightcast data) is unique to Australia. In other countries, this sector is reasonably represented (differences of less than one percentage point) or underrepresented.

The results for Canada are found in Figure 5.4. Finance, Insurance and Real Estate (FIRE) is most overrepresented sector (+8 percentage points), followed by Accommodation, Food, Arts and Recreation (+3 percentage points). The least represented are Professional, Scientific, Technical, Administrative and Other Services and Construction (both -4 percentage points) followed by Education and Health (-3 pp) and Natural Resources and Agriculture (-1pp). The differences for other sectors are under 1pp.

Figure 5.4. Comparison of distributions across sectors in 2022, Canada

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.

Note: Data for Q1 2022; sectors are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented, FIRE stands for Finance, Insurance and Real Estate.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey, 2022) data.

In the UK (Figure 5.5), Education and Health are the most overrepresented sector (+12pp). The differences between shares in the official and the Lightcast data for other overrepresented sectors is considerably smaller (+3pp for Manufacturing and Utilities; +1pp for Public Administration and Safety and FIRE). The most underrepresented sectors are Accommodation, Food, Arts and Recreation (-8pp) followed by Trades, Transport and Warehousing (-6pp), Information Media and Telecommunications (-2pp) and Professional, Scientific, Technical, Administrative and Other Services (-1pp). The differences in shares for the two remaining sectors (Construction and Natural Resources) are under 1 percentage point.

Results for the US are found in Figure 5.6. The differences in distributions range from +7 percentage points (overrepresentation) for Education, Healthcare and Social Assistance to -10 percentage points (underrepresentation) for Professional, Scientific, Technical, Administrative and Other Services. The overrepresented sectors include FIRE (+6pp) followed by Manufacturing (+2pp). Information; Transportation, Warehousing and Utilities; Wholesale and Retail Trade are well-represented (the difference is smaller than 1pp). Among underrepresented sectors are Government (-9pp), Accommodation, Food, Arts and Recreation (-6pp) and Construction (-2pp).

Figure 5.5. Comparison of distributions across sectors in 2022, the UK

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.

Note: Data for Q2 2022; sectors are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented, FIRE stands for Finance, Insurance and Real Estate.
Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series, 2022) data.

Figure 5.6. Comparison of distributions across sectors in 2022, the US

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
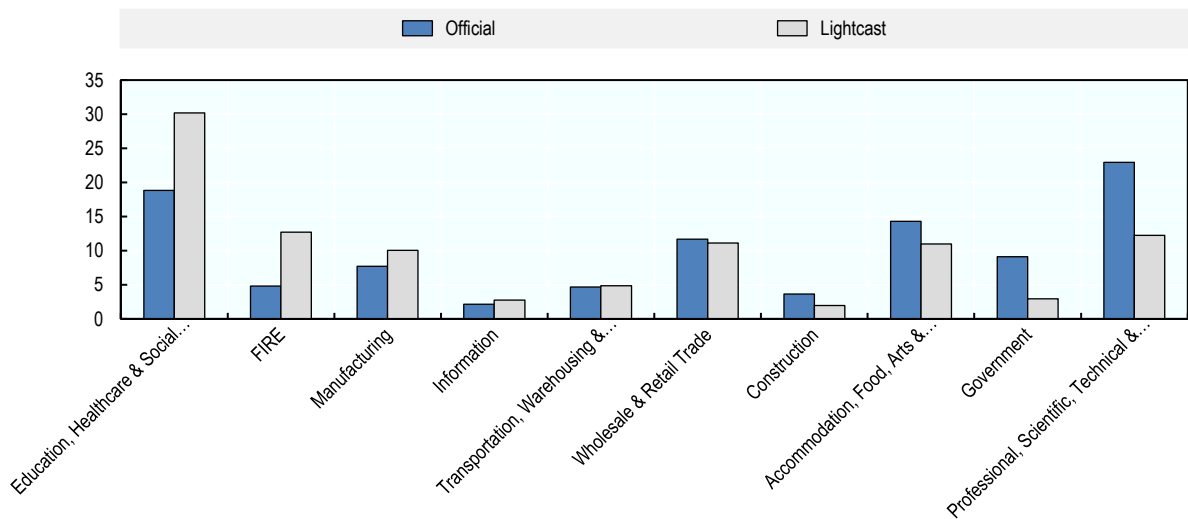
Note: Data for Q2 2022; sectors are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented, FIRE stands for Finance, Insurance and Real Estate.
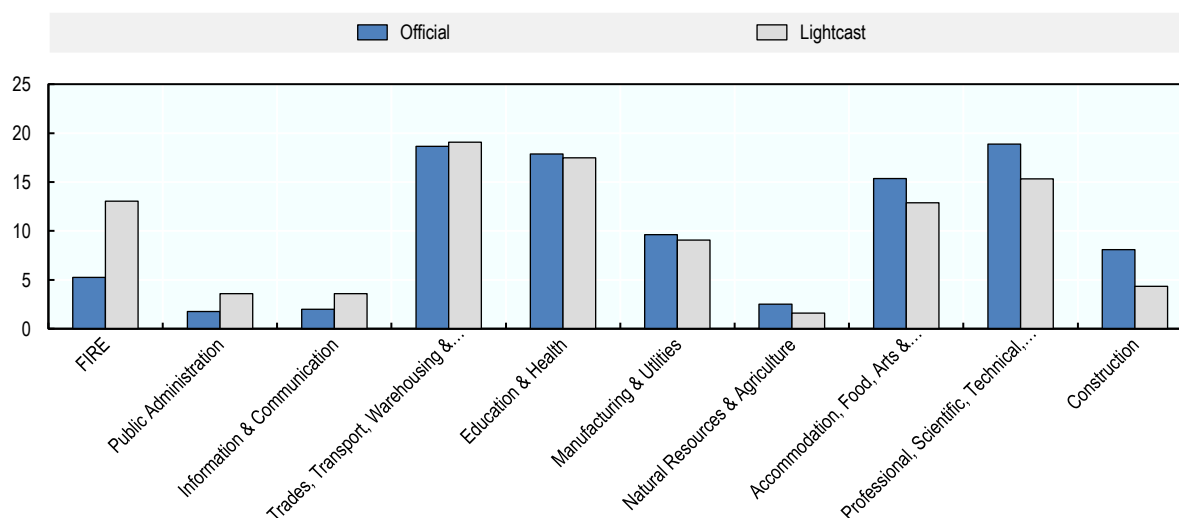Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey, 2022) data.

Results over time

This subsection presents an assessment of the representativeness of the Lightcast data by industrial sector between 2015 and 2022 using quarterly intervals. Overall, Figure 5.7 - Figure 5.10 show relatively stable patterns over time with some exceptions.

In Australia, out of ten industrial sectors shown in Figure 5.7 four are consistently overrepresented, four are consistently underrepresented and representation of the remaining two differs across periods. Two of the most overrepresented sectors are Education and Health (15 percentage point difference between shares in the Lightcast and the official data on average) and Public Administration and Safety (11 percentage point difference on average). Over time, the discrepancy for the Public Administration and Safety decreased from 11 to 6 percentage points. The most underrepresented sector is Professional, Scientific, Technical, Administrative and Other Services with the differences in shares of -16 percentage points on average. Yet, the underrepresentation of this sector in the Lightcast data decreased from -15pp to -9pp. Representativeness of other sectors is generally stable over time with the exception of Accommodation, Food, Arts & Recreation; Natural Resources and Trades, Transport and Warehousing. On average over time, annual aggregations (not reported for brevity) do not necessarily offer a closer match to the official data. In four out of ten sectors, the differences between the Lightcast and the official data in quarterly aggregations are smaller than in annual ones.

Figure 5.7. Representativeness by industrial sectors over time, Australia

Differences between vacancy shares reported by the Lightcast data and the official data.
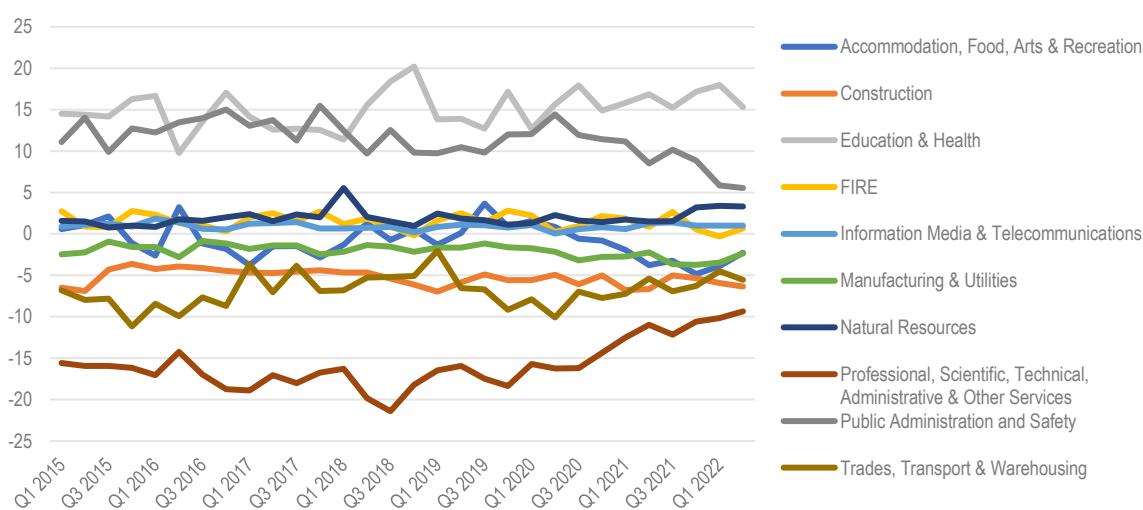
Note: FIRE stands for Finance, Insurance and Real Estate.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Table 5.2. Summary statistics for differences in shares of industrial sectors, Australia

Sector	Mean	Std. dev.	Min	Max
Accommodation, Food, Arts & Recreation	-0.84	2.12	-4.83	3.67
Construction	-5.27	0.95	-6.98	-3.62
Education & Health	15.04	2.30	9.78	20.19
FIRE	1.49	0.92	-0.29	2.82
Information, Media & Telecommunications	0.94	0.38	0.03	1.84
Manufacturing & Utilities	-2.07	0.79	-3.74	-0.83
Natural Resources	1.92	0.96	0.77	5.53
Professional, Scientific, Technical, Administrative & Other Services	-15.78	2.91	-21.39	-9.35
Public Administration & Safety	11.41	2.39	5.56	15.48
Trades, Transport & Warehousing	-6.84	1.99	-11.18	-2.04

Note: The table shows summary statistics for each of the series shown in Figure 5.7.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

The Lightcast data for Canada show a better representation overall with the differences generally staying within 5 percentage points for both underrepresentation and overrepresentation (Figure 5.8). The most notable exception is strongly overrepresented Finance, Insurance and Real Estate with the share of vacancies in the Lightcast data exceeding the official data by 7 percentage points on average. Moreover, this overrepresentation was uneven over time with an overall upward trend. The most underrepresented sectors on average are Accommodation, Food, Arts and Recreation (-4pp) and Construction (-3pp), although there is considerable variation from quarter to quarter, as changes over time for the same sector are larger compared to Australia. In Canada, alternative intervals (monthly or annually) should also be considered as both of them offer closer match to the distributions observed in the official data (Annex F shows monthly aggregations, while the annual ones are not reported for brevity).

Figure 5.8. Representativeness by industrial sectors over time, Canada

Differences between vacancy shares reported by the Lightcast data and the official data.
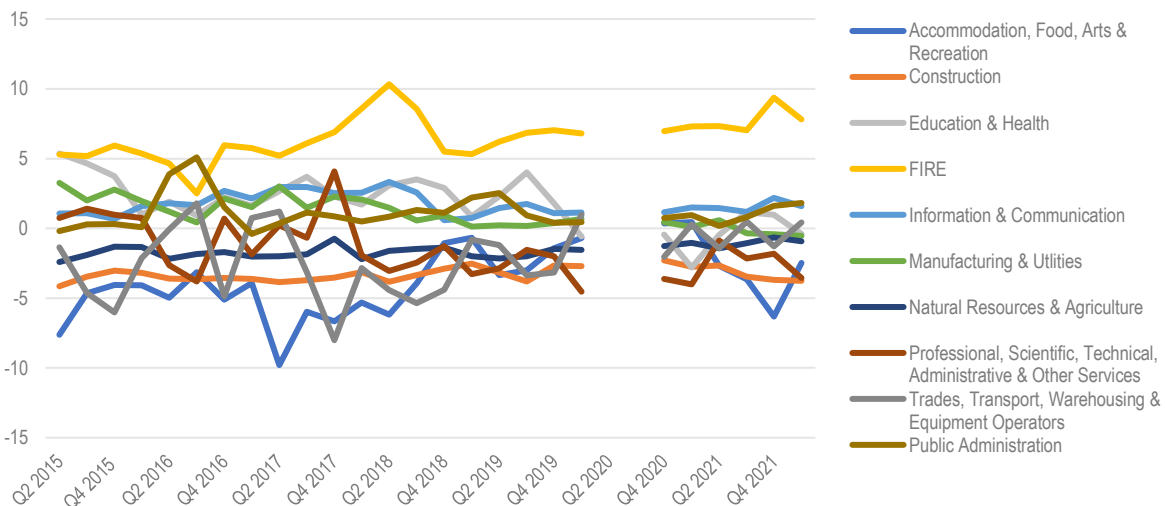
Note: FIRE stands for Finance, Insurance and Real Estate.
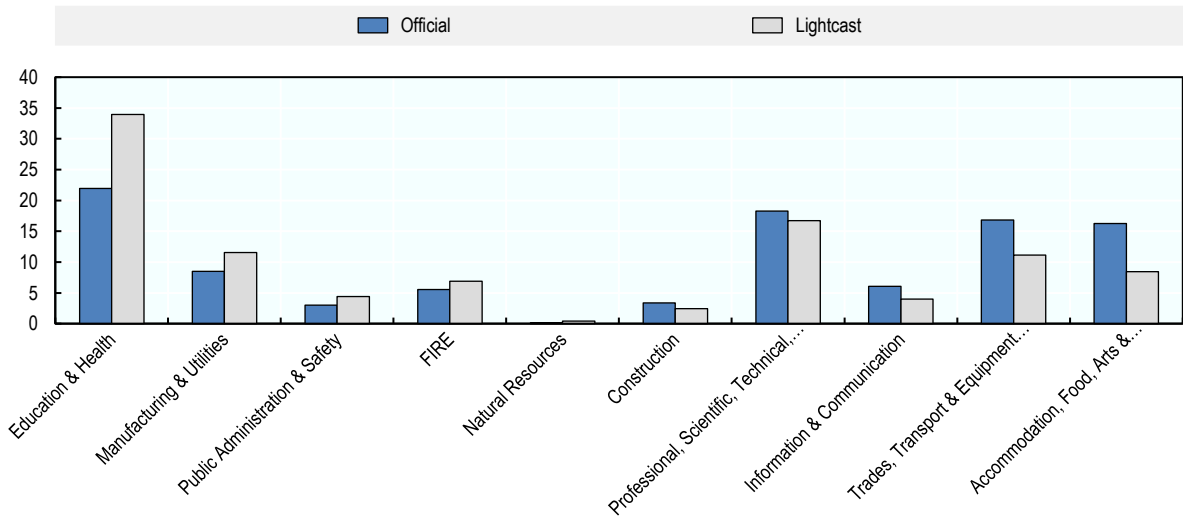Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Table 5.3. Summary statistics for differences in shares of industrial sectors, Canada

Sector	Mean	Std. dev.	Min	Max
Accommodation, Food, Arts & Recreation	-3.84	2.48	-9.79	0.44
Construction	-3.31	0.49	-4.15	-2.30
Education & Health	1.84	1.85	-2.80	5.37
FIRE	6.54	1.61	2.50	10.34
Information & Communication	1.75	0.77	0.58	3.32
Manufacturing & Utilities	1.10	1.09	-0.53	3.26
Natural Resources & Agriculture	-1.59	0.48	-2.41	-0.64
Professional, Scientific, Technical, Administrative & Other Services	-1.50	2.05	-4.54	4.10
Trades, Transport, Warehousing & Equipment	-2.10	2.54	-8.02	1.81
Public Administration	1.13	1.22	-0.40	5.10

Note: The table shows summary statistics for each of the series shown in Figure 5.8.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

In the UK (Figure 5.9), the differences in shares between the Lightcast and the official data tend to be no larger than 5 percentage points. Exceptions include Education and Health (+9pp differences on average over time and increasing) and Trades, Transport and Equipment Operators (-10pp differences on average and increasing in absolute value). The patterns are quite stable over time with the exception of the COVID-19 period, during which differences decreased for all sectors except for Education and Health, while the rankings of the two most underrepresented sectors (Trades, Transport and Equipment; Accommodation, Food, Arts and Recreation) flipped. Annual aggregations instead of the quarterly ones do not offer much improvement in the match between distributions in the two data sets as the mean difference is marginally lower (by less then 0.1) in only three sectors out of ten (not reported for brevity).

Figure 5.9. Representativeness by industrial sectors over time, the UK

Differences between vacancy shares reported by the Lightcast data and the official data.
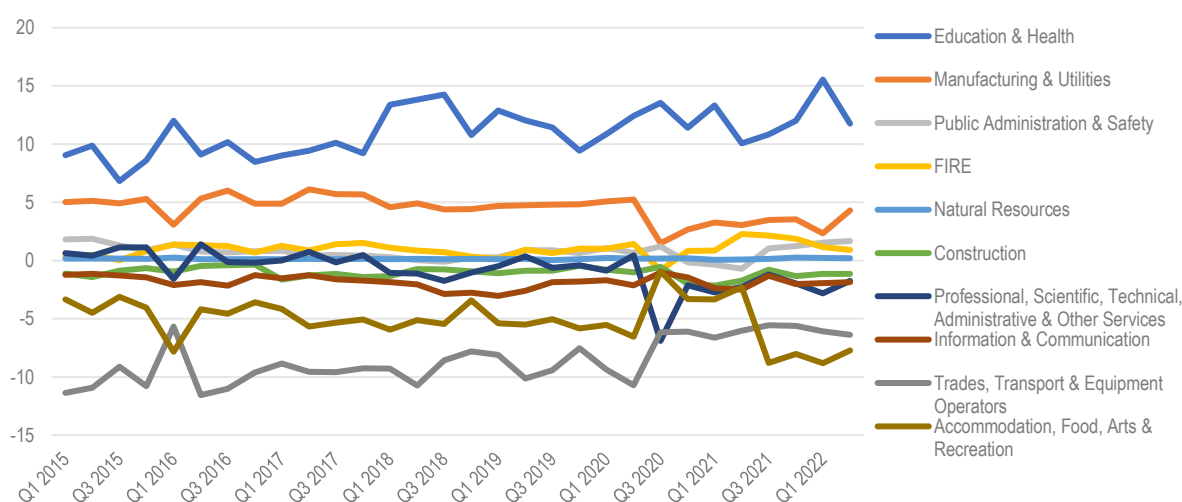
Note: FIRE stands for Finance, Insurance and Real Estate.
Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series) data.

Table 5.4. Summary statistics for differences in shares of industrial sectors, the UK

| Sector	Mean	Std. dev.	Min	Max
Education & Health	9.27	1.89	6.82	12.00
Manufacturing & Utilities	4.69	0.91	3.09	5.30
Public Administration & Safety	1.45	0.41	0.89	1.88
FIRE	0.61	0.50	0.04	1.36
Natural Resources	0.18	0.04	0.16	0.26
Construction	-1.00	0.28	-1.41	-0.66
Professional, Scientific, Technical, Administrative & Other Services	0.35	1.11	-1.57	1.13
Information & Communication	-1.44	0.37	-2.10	-1.14
Trades, Transport & Equipment Operators	-9.57	2.34	-11.36	-5.66
Accommodation, Food, Arts & Recreation	-4.56	1.90	-7.82	-3.11

Note: The table shows summary statistics for each of the series shown in Figure 5.9.
Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series) data.

In the US (Figure 5.10), the differences in vacancy distribution with the official data range between +11 (Education, Healthcare and Social Assistance) and -10 (Professional, Scientific, Technical and Other Services) on average. Three sectors are consistently overrepresented (Information, FIRE; Education, Healthcare and Social Assistance). Four sectors are consistently underrepresented (Government; Construction; Professional, Scientific, Technical and Other Services; Accommodation, Food Arts and Recreation). The patterns over time appear stable except for Accommodation, Food, Arts and Recreation where underrepresentation is decreasing and Transportation, Warehousing and Utilities, which went from an overrepresented to a well-represented sector. Noteworthy, underrepresentation of Government tends to somewhat increase and so does overrepresentation of FIRE. Quarterly aggregations offer the closest fit on average. Alternative aggregations (monthly and annually) only show a marginal decrease in the differences in distributions in four sectors out of ten (Annex F shows monthly aggregations while the annual ones are not reported for brevity).

Figure 5.10. Representativeness by industrial sectors over time, the US

Differences between vacancy shares reported by the Lightcast data and the official data.

Note: FIRE stands for Finance, Insurance and Real Estate.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

Table 5.5. Summary statistics for differences in shares of industrial sectors, the US

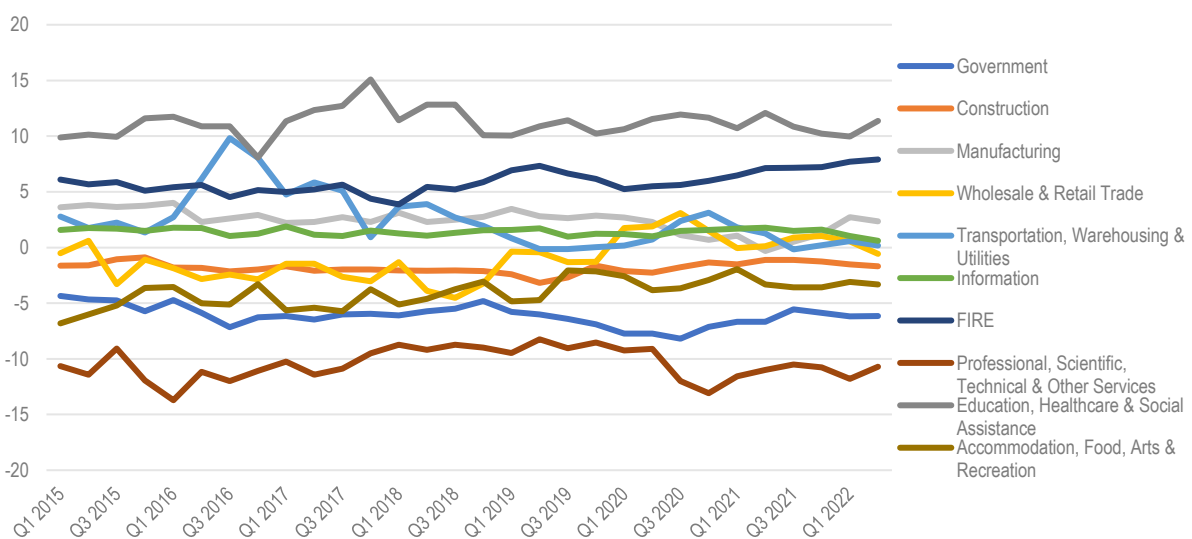| Sector	Mean	Std. dev.	Min	Max
Government	-6.11	0.93	-8.18	-4.35
Construction	-1.82	0.49	-3.18	-0.88
Manufacturing	2.43	1.03	-0.31	4.00
Wholesale & Retail Trade	-0.97	1.89	-4.54	3.10
Transportation, Warehousing & Utilities	2.48	2.52	-0.17	9.83
Information	1.40	0.32	0.62	1.90
FIRE	5.90	1.00	3.87	7.91
Professional, Scientific, Technical & Other Services	-10.46	1.42	-13.70	-8.26
Education, Healthcare & Social Assistance	11.18	1.28	8.08	15.09
Accommodation, Food, Arts & Recreation	-4.04	1.25	-6.81	-1.94

Note: The table shows summary statistics for each of the series shown in Figure 5.10.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

By region

This subsection assesses representativeness of the Lightcast data against the official data sources along the geographical dimension. Among the four countries considered in this report, all except the UK report regional vacancy data. Analysis below first presents the 2022 comparisons of distributions in job openings by TL2 region and then shows trends over time using quarterly aggregations.

Unlike the representativeness by occupation or industry, the regional representativeness of the Lightcast data has not been assessed in the literature in an international perspective. The geographical dimension of the scraped vacancy data (information on the location of a job opening) is undoubtedly an important attribute that makes this data set well-suited for a wide range of regional analysis applications seeking to understand and explain spatial organisation of economic activities and other processes in space.

Results for 2022

In the three countries with available data, regional representativeness of the Lightcast vacancy data is quite good. Most regions (5 out of 8 in Australia, 7 out of 13 in Canada and 46 out of 50 in the US) are well-represented in 2022 (one percentage point or less in differences between the Lightcast and officially reported shares across regions).

- In Australia (Figure 5.11), the most overrepresented region in 2022 is New South Wales where the Lightcast data reports 34% of all vacancies compared to 32% reported by the Australian Bureau of Statistics. The most underrepresented region is Western Australia with 4pp difference between the officially reported share of vacancies and one reported by Lightcast.

- The differences in the shares of vacancies reported by the Lightcast data and the official data in Canada (Figure 5.12) are comparable to that of Australia (4pp or lower), except for a very large (13pp) difference in Quebec. Potentially the Lightcast algorithm is not well-suited to scrape announcements in French, which is the official language in the province. Larger regional economies are predictably overrepresented in the Lightcast data (e.g. Ontario, British Columbia and Alberta) but so are smaller regions such as Saskatchewan and Northwest Territories, for which differences in shares stand at 4pp and 3pp, respectively. Besides Quebec, the only region with underrepresentation exceeding 1pp is Nova Scotia, where the difference is 2pp.

- In the US, the regional representativeness of the Lightcast data is very high (Figure 5.13). Only in California differences between the Lightcast and officially reported shares differ by 2 percentage

points. In all other states and the District of Columbia, the differences are less than 1 percentage point.

Figure 5.11. Comparison of distributions across states and territories in 2022, Australia

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
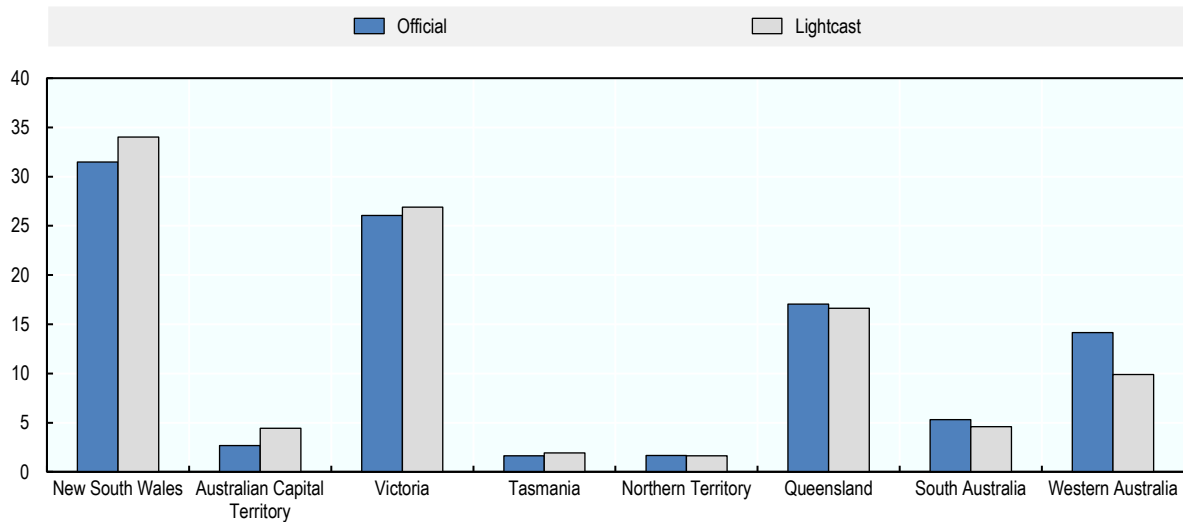
Note: Data for 21 April – 20 May 2022 in the Lightcast data, Q2 2022 in the official data; regions are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly, 2022) data.

Figure 5.12. Comparison of distributions across provinces and territories in 2022, Canada

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
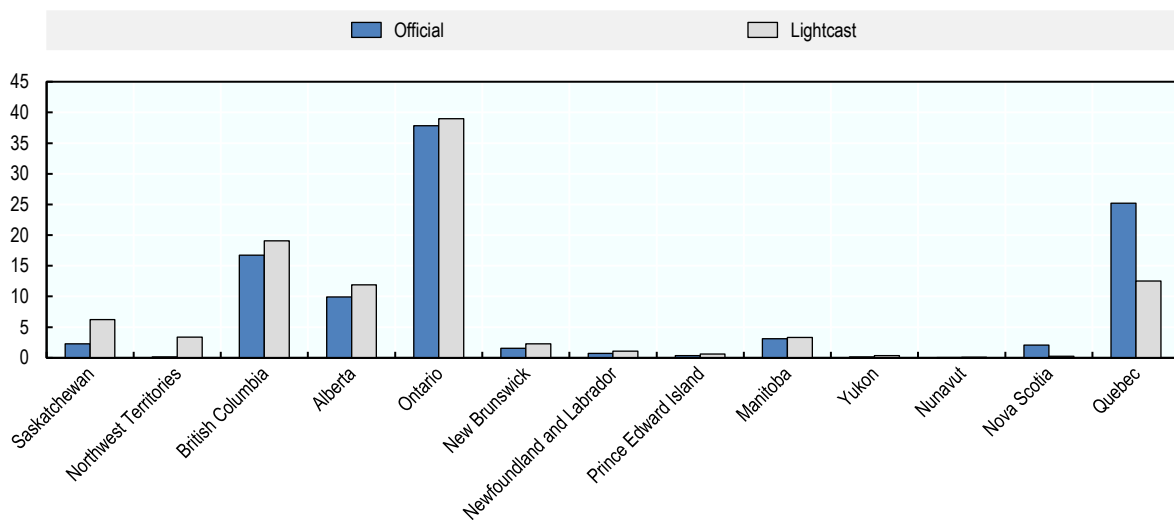
Note: Data for Q1 2022; regions are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey, 2022) data.

Figure 5.13. Comparison of distributions across states and the District of Columbia in 2022, the US

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
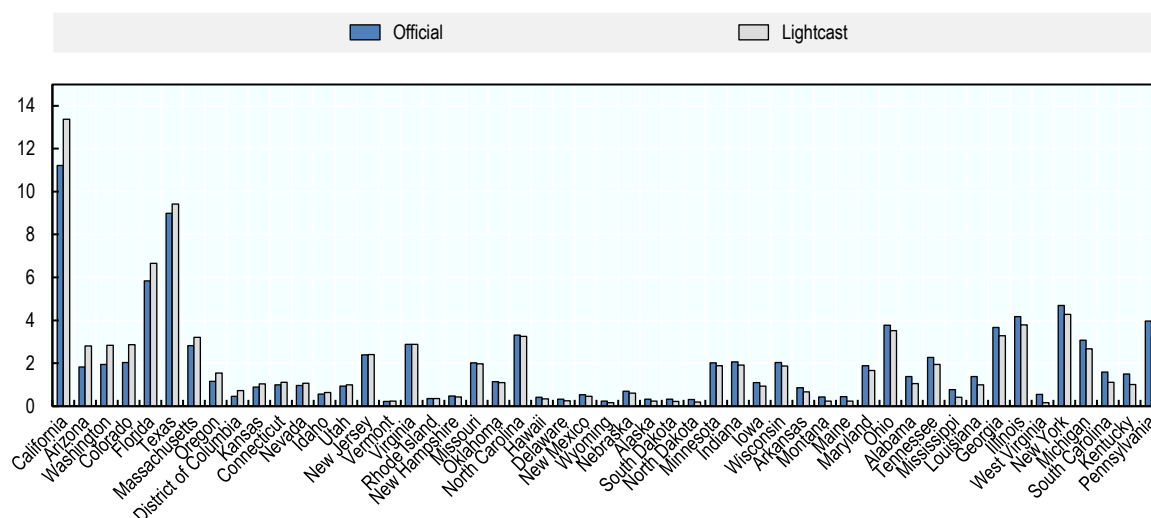
Note: Data for Q2 2022; regions are ordered by the difference in shares from the most overrepresented in the Lightcast data to the most underrepresented.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey, 2022) data.

Over time

Differences in shares across TL2 regions between the Lightcast and the official data tend to be quite small except for specific regions (Figure 5.14-Figure 5.16).[22] Most of the regions are underrepresented in some periods and overrepresented in others.

In Australia (Figure 5.14), the differences between the shares in the Lightcast and the official data range from -4 to 3 percentage points on average. Despite the relatively small range, the magnitude of the differences and the rankings of the regions from the most to the least represented strongly depend on the period in focus. There are no clearly observable trends during the period Q1 2015 – Q2 2022, except for the Western Australia region where underrepresentation generally increases and for Victoria, which went from underrepresentation to overrepresentation. Annual aggregations offer slightly closer match between distributions in the Lightcast and the official data in four regions out of eight (not reported for brevity).

---

[22] Sensitivity of regional analysis results in countries with regions-outliers can be checked by, for example, repeating analysis after excluding the least represented regions or regions where the changes over time are the largest.

Figure 5.14. Representativeness by states and territories over time, Australia

Differences between vacancy shares reported by the Lightcast data and the official data.
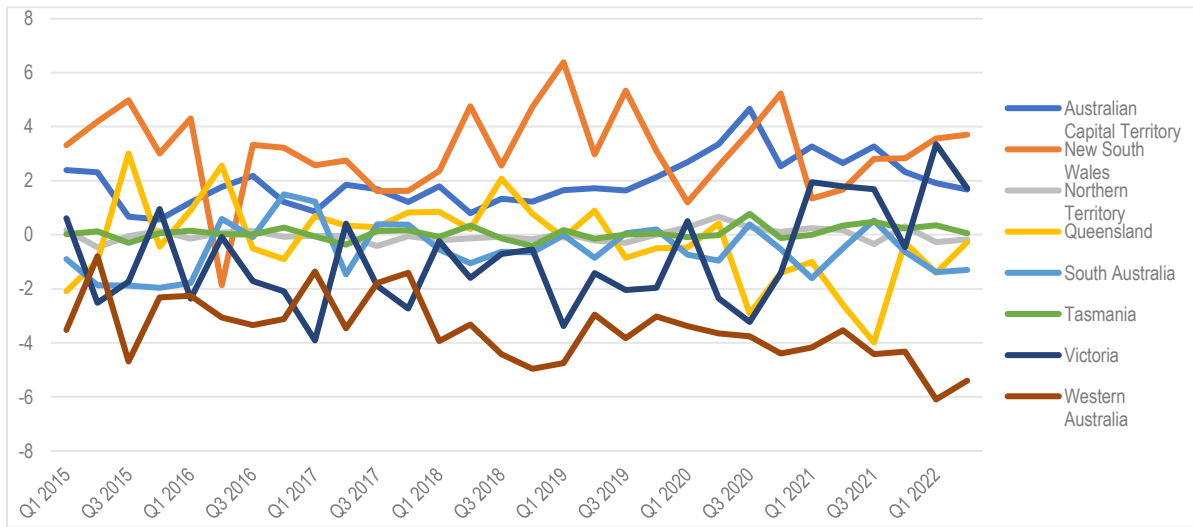
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Table 5.6. Summary statistics for differences in shares by states and territories, Australia

Region	Mean	Std. dev.	Min	Max
Australian Capital Territory	1.95	0.91	0.56	4.66
New South Wales	3.13	1.58	-1.87	6.38
Northern Territory	-0.02	0.25	-0.47	0.67
Queensland	-0.23	1.52	-3.98	3.01
South Australia	-0.54	0.92	-1.97	1.49
Tasmania	0.07	0.25	-0.42	0.77
Victoria	-0.85	1.82	-3.91	3.35
Western Australia	-3.52	1.21	-6.10	-0.79

Note: The table shows summary statistics for each of the series shown in Figure 5.14.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Changes in spatial representativeness over time in Canada are very region-specific (Figure 5.15). A group of small regions (Manitoba, New Brunswick, Newfoundland and Labrador, Nunavut, Prince Edward Island and Yukon) are consistently well-represented in the data in all years considered. Representativeness of larger regions is more volatile. The differences in shares reported by the Lightcast and the official data are the largest in Quebec, which went from overrepresentation of 7pp in 2016 to underrepresentation of 15pp in the last quarter of 2021. Another region with highly variable representation, Ontario, is underrepresented in some periods and overrepresented in others with magnitude of the differences ranging from -6pp to +5pp. Monthly and annual aggregations do not considerably improve the match in distributions, as on average, they offer smaller differences in two and five regions out of 13, respectively.

Figure 5.15. Representativeness by provinces and territories over time, Canada

Differences between vacancy shares reported by the Lightcast data and the official data.
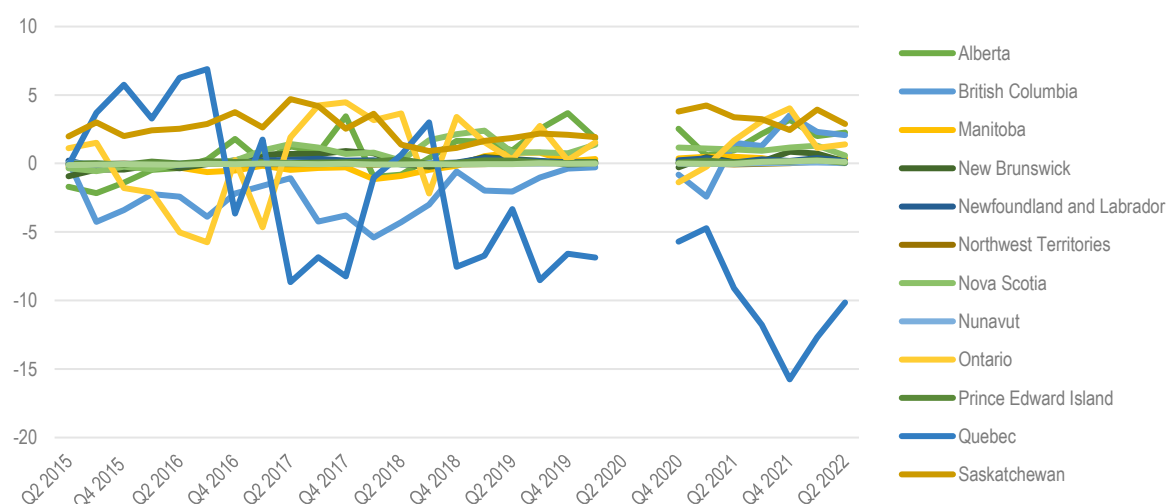
Notes: Data for Q2 and Q3 in 2020 are missing due to the COVID-19 pandemic.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Table 5.7. Summary statistics for differences in shares by provinces and territories, Canada

Region	Mean	Std. dev.	Min	Max
Alberta	0.97	1.58	-2.17	3.68
British Columbia	-1.50	2.27	-5.40	3.52
Manitoba	-0.05	0.50	-1.13	0.80
New Brunswick	0.16	0.46	-0.94	0.90
Newfoundland and Labrador	0.13	0.18	-0.25	0.40
Northwest Territories	-0.02	0.05	-0.09	0.13
Nova Scotia	0.78	0.74	-0.56	2.39
Nunavut	0.01	0.03	-0.03	0.06
Ontario	0.67	2.82	-5.74	4.47
Prince Edward Island	0.10	0.09	-0.09	0.30
Quebec	-3.96	6.21	-15.76	6.89
Saskatchewan	2.71	0.99	0.90	4.69
Yukon	-0.01	0.08	-0.11	0.21

Note: The table shows summary statistics for each of the series shown in Figure 5.15.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Spatial representativeness of the Lightcast data in the US is very high (Figure 5.16). On average, the differences between shares in the Lightcast and the official data exceed 0.5 percentage points only in seven states (Arizona, California, Colorado and Massachusetts are overrepresented, whilst New York, Pennsylvania and South Carolina are underrepresented). At the extremes, overrepresentation reaches 11 percentage points in California and 8 percentage points in Arizona; underrepresentation reaches 3 percentage points in Georgia, Illinois, New York, Ohio, and Pennsylvania. Arizona is the only state that is consistently overrepresented in all periods, while Louisiana, South Carolina and West Virginia are consistently underrepresented. As in other countries, alternative aggregations of the regional data (monthly and annually) do not improve the match between distributions in the Lightcast and the official data. On

average, the differences are slightly smaller in 15 states for monthly aggregations and in 22 states for annual ones out of total 51 states and the District of Columbia.

Figure 5.16. Representativeness by states and the District of Columbia over time, the US

Differences between vacancy shares reported by the Lightcast data and the official data.
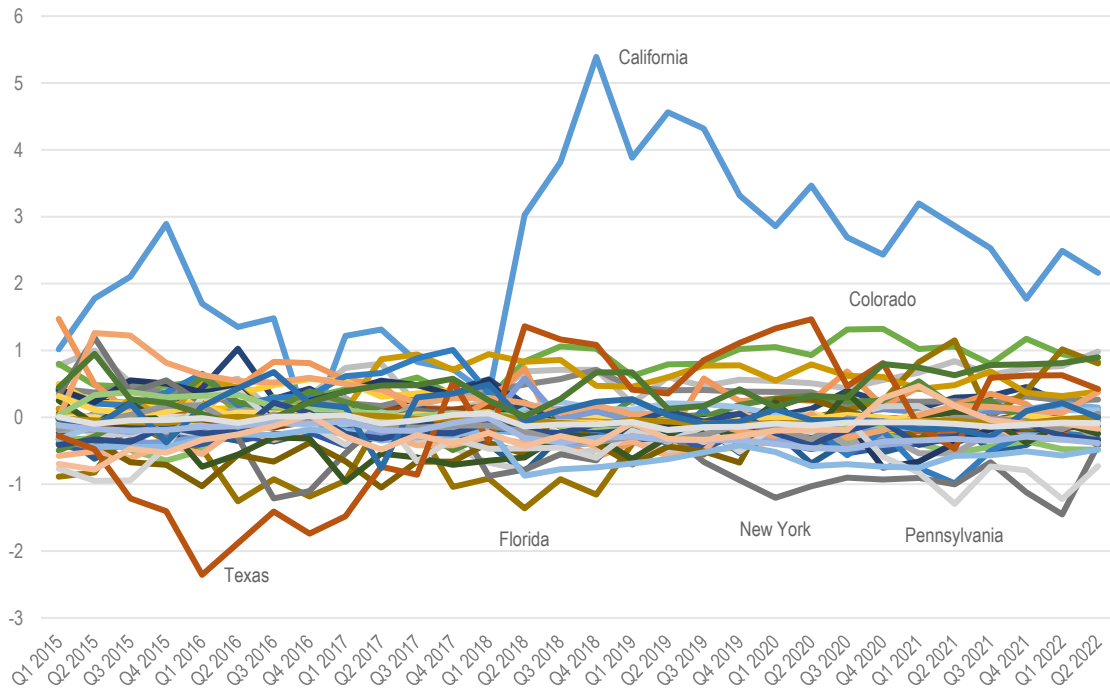
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

Table 5.8. Summary statistics for differences in shares by states and District of Columbia, the US

State	Mean	Std. dev.	Min	Max
Alabama	-0.08	0.31	-0.47	1.18
Alaska	-0.04	0.09	-0.33	0.12
Arizona	0.82	1.43	0.07	8.42
Arkansas	-0.15	0.21	-0.58	0.26
California	2.65	1.98	-0.23	10.75
Colorado	0.65	0.60	-2.05	1.32
Connecticut	0.20	0.31	-0.98	1.03
Delaware	0.01	0.11	-0.33	0.22
District of Columbia	0.32	0.18	-0.39	0.71
Florida	-0.16	1.07	-1.36	4.22
Georgia	-0.38	0.52	-2.95	0.21
Hawaii	-0.09	0.08	-0.45	0.02
Idaho	-0.08	0.14	-0.57	0.14
Illinois	0.12	0.80	-3.45	1.47
Indiana	-0.27	0.38	-1.92	0.17
Iowa	0.11	0.25	-0.28	0.57
Kansas	0.06	0.55	-0.32	2.95
Kentucky	-0.28	0.32	-1.56	0.30
Louisiana	-0.35	0.21	-1.32	-0.15

Maine	-0.14	0.11	-0.46	0.07
Maryland	-0.11	0.41	-1.86	0.47
Massachusetts	0.56	0.42	-0.16	2.19
Michigan	-0.05	0.61	-2.08	1.01
Minnesota	0.02	0.44	-1.91	0.60
Mississippi	-0.20	0.25	-0.47	0.89
Missouri	-0.23	0.41	-0.61	1.76
Montana	-0.12	0.10	-0.44	0.06
Nebraska	-0.03	0.12	-0.62	0.08
Nevada	-0.05	0.23	-1.03	0.23
New Hampshire	-0.06	0.10	-0.46	0.08
New Jersey	-0.01	0.49	-1.87	0.56
New Mexico	-0.10	0.11	-0.60	0.03
New York	-0.58	0.80	-3.33	1.19
North Carolina	-0.46	0.57	-2.93	0.27
North Dakota	-0.08	0.09	-0.25	0.22
Ohio	-0.35	0.56	-2.82	0.37
Oklahoma	-0.17	0.26	-1.16	0.58
Oregon	0.46	1.20	-0.30	6.59
Pennsylvania	-0.55	0.66	-3.39	0.41
Rhode Island	0.09	0.46	-0.09	2.50
South Carolina	-0.52	0.27	-1.54	-0.16
South Dakota	0.05	0.26	-0.19	1.10
Tennessee	-0.11	0.44	-0.41	2.10
Texas	0.04	1.18	-2.35	2.70
Utah	-0.10	0.16	-0.37	0.34
Vermont	-0.05	0.05	-0.23	0.05
Virginia	0.01	0.38	-1.41	0.68
Washington	0.42	0.30	-0.13	0.95
West Virginia	-0.26	0.13	-0.53	-0.01
Wisconsin	-0.28	0.38	-1.76	0.47
Wyoming	-0.07	0.07	-0.23	0.06

Note: The table shows summary statistics for each of the series shown in Figure 5.16.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

6 Conclusion

Data on online job postings represents an important data source for labour market analysis. Online job postings can provide a richer context of labour demand relative to vacancy surveys. For instance, detailed and timely information on labour demand across regions, sectors, occupations and over time are contained in job postings. However, analysis using this data often also requires that data on online job postings provide a fair and accurate representation of labour demand.

This paper compares the distribution of Lightcast data against those of official national surveys on vacancies and job openings for 2015-2022. Online job postings may not reflect the entire labour market. The data is limited to those posted online, and online hiring may differ across sectors, occupations and even regions within the same country. Moreover, single online job postings may be used to hire multiple people. Official national surveys on vacancies and job openings are already conducted and can be used to assess the representativeness of Lightcast data.

This paper shows that benchmarking Lightcast data against official data on vacancy counts is feasible and provides useful information on differences in the distribution between the two types of sources. The benchmarking exercise presented in this paper is based on publicly available data from official statistical sources. While official vacancy data is not as rich in information as individual job postings, the surveys are designed to reflect labour demand as accurately as possible and are therefore an informative benchmark against which to assess Lightcast data.

Overall, Lightcast data closely tracks the distribution observed in official vacancy data, but differences remain. While total counts of vacancies in Lightcast are generally below those in official data, the relative shares are quite similar and stable over time. Some sectors and regions are consistently over- or underrepresented in Lightcast relative to official statistical sources. For instance, Lightcast data tend to overrepresent the Education and Health sector in Australia, the UK and the US by 7 to 15 percentage points.

Analysts that use Lightcast, or similar data from other providers, can consider performing similar benchmarking for their analyses and applying statistical techniques to align distributions where necessary. Benchmarking Lightcast data with official vacancy data can help ascertain whether research findings are sensitive to variation in the data gathering and information extraction of online job postings. Where appropriate, statistical techniques, such as data weighting, could be used to correct for under- and overrepresentation of some dimensions of online job postings data.

References

Autor, D., F. Levy and R. Murnane (2003), "The skill-content of recent technological change: An empirical investigation", *Quarterly Journal of Economics*, Vol. 118/4, pp. 1279-1333, https://academic.oup.com/qje/article-abstract/118/4/1279/1925105. [1]

Bacolod, M., B. Blum and W. Strange (2009), "Skills in the city", *Journal of Urban Economics*, Vol. 65/2, pp. 136-153, https://doi.org/10.1016/J.JUE.2008.09.003. [3]

Cammeraat, E. and M. Squicciarini (2021), "Burning Glass Technologies' data use in policy-relevant analysis: An occupation-level assessment"*, OECD Science, Technology and Industry Working Papers*, No. 2021/05, OECD Publishing, Paris, https://doi.org/10.1787/cd75c3e7-en. [12]

Carnevale, A., T. Jayasundera and D. Repnikov (2014), *Understanding Online Job Ads Data: A Technical Report*, Washington, DC: Georgetown University Center on Education and the Workforce. [9]

Duranton, G. and D. Puga (2005), "From sectoral to functional urban specialisation", *Journal of Urban Economics*, Vol. 57/2, pp. 343-370, https://doi.org/10.1016/j.jue.2004.12.002. [4]

Feser, E. (2003), "What regions do rather than make: a Proposed set of knowledge-based occupation clusters", *Urban Studies*, Vol. 40/10, pp. 1937-1958, https://doi.org/10.1080/0042098032000116059. [2]

Hershbein, B. and L. Kahn (2018), "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings", *American Economic Review*, Vol. 108/7, pp. 1737-1772, https://doi.org/10.1257/aer.20161570. [5]

Lancaster, V., D. Mahoney-Nair and N. Ratcliff (2019), *Review of Burning Glass Job-ad Data*, University of Virginia, Charlottesville, VA, https://biocomplexity.virginia.edu/sites/default/files/projects/Technical%20Report%20Review%20of%20BGT%20Job-ad%20Data.pdf (accessed on 24 September 2022). [10]

Mohnen, P., E. Berkes and B. Taska (2018), "The Effect of Skill Mismatch on Early Career Outcomes of College Graduates: Evidence from Online Job Postings", in *Essays in Labor Economics*, Northwestern University, Ewanston, IL, https://search.proquest.com/openview/2df7abf37ef24942180fe75e26fb6928/1?pq-origsite=gscholar&cbl=18750 (accessed on 26 September 2022). [11]

Morgan, P. et al. (2017), "Scarcity of Primary Care Positions May Divert Physician Assistants into Specialty Practice", *Medical Care Research and Review*, Vol. 74/1, pp. 109-122, https://doi.org/10.1177/1077558715627555/ASSET/IMAGES/LARGE/10.1177_10775587156 27555-FIG2.JPEG. [8]

Sigelman, M. et al. (2019), *How New Skills Are Rewriting the DNA of the Job Market*, Burning Glass Technologies, https://www.emarketer.com/content/mobile-time-spent-2018 (accessed on 25 March 2020). [6]

Vermeulen, W. and F. Gutierrez Amaros (2024), "How well do online job postings match national sources in European countries?: Benchmarking Lightcast data against statistical and labour agency sources across regions, sectors and occupation", *OECD Local Economic and Employment Development (LEED) Papers*, No. 2024/02, OECD Publishing, Paris, https://doi.org/10.1787/e1026d81-en. [7]

Annex A. A comparison of correlations between employment and vacancies in Canada by occupation, official data

Table A A.1. Pairwise correlations between official employment data and vacancy data, Canada

Occupational group	Counts	Shares
Total	0.84*	
Management	0.69*	-0.06
Business, finance and administration	0.95*	-0.03
Natural and applied sciences and related	0.95*	0.29
Health	0.89*	0.82*
Education, law and social, community and government services	0.74*	0.46*
Art, culture, recreation and sport	0.19	-0.01
Sales and service	-0.14	0.54*
Trades, transport and equipment operators and related	0.46*	-0.10
Natural resources, agriculture and related production	0.32	0.17
Manufacturing and utilities	-0.38	-0.33

Note: National quarterly data for Q1 2015 - Q4 2021 (Q1 and Q2 2020 are missing); total number of observations is 29; * significant at the 0.95 level.
Source: OECD calculations based on Statistics Canada data (Tables 14-10-0356-01 and 14-10-0296-01).

Annex B. Ratios of the number of vacancies in Lightcast to official data

Figure A B.1. Ratio of the Lightcast to officially reported vacancies by occupation, quarterly aggregations, Canada
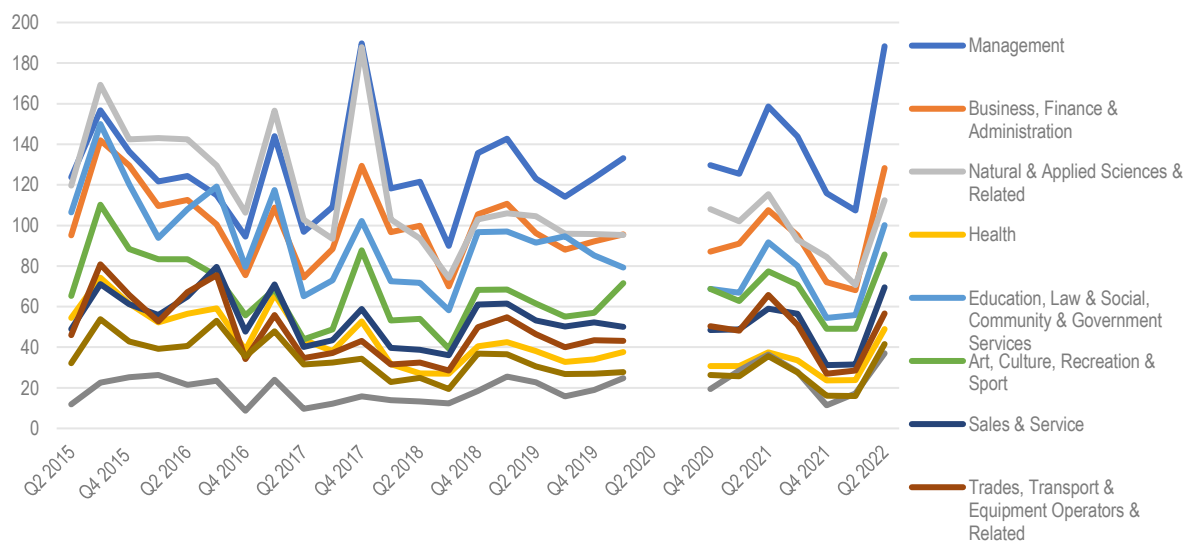
Note: Data for Q2 and Q3 in 2020 are missing due to the COVID-19 pandemic.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Table A B.1. Summary statistics for the ratio of vacancies by occupation, quarterly aggregations, Canada

Occupation	Mean	Std. dev.	Min	Max	Coef. of var.
Management	129.01	24.12	89.98	189.67	0.19
Business, Finance & Administration	98.87	19.03	68.14	141.92	0.19
Natural & Applied Sciences & Related	112.98	28.03	70.57	187.80	0.25
Health	42.12	13.45	23.70	74.17	0.32
Education, Law & Social, Community & Government Services	88.89	22.71	54.39	149.99	0.26
Art, Culture, Recreation & Sport	66.77	16.30	39.44	110.19	0.24
Sales & Service	52.98	12.45	31.15	79.57	0.23
Trades, Transport & Equipment Operators & Related	47.80	14.40	27.01	80.79	0.30
Natural Resources, Agriculture & Related Production	20.17	7.55	8.79	37.01	0.37
Manufacturing & Utilities	32.76	9.85	15.95	53.76	0.30

Note: The table shows summary statistics and coefficient of variation for each of the series shown in Figure A B.1.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Figure A B.2. Ratio of the Lightcast to officially reported vacancies by unaggregated sector, quarterly aggregations, Australia

Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Table A B.2. Summary statistics for the ratio of vacancies by unaggregated sector, quarterly aggregations, Australia

Sector	Mean	Std. dev.	Min	Max	Coef. of var.
Mining	41.73	15.05	27.10	89.65	0.36
Manufacturing	13.43	5.76	6.62	39.91	0.43
Electricity, Gas, Water and Waste Services	35.30	14.74	18.44	88.00	0.42
Construction	6.25	3.40	2.75	21.67	0.54
Wholesale Trade	7.20	3.07	3.11	17.26	0.43
Retail Trade	16.89	4.45	12.22	35.96	0.26
Accommodation and Food Services	20.28	9.06	9.39	54.23	0.45
Transport, Postal and Warehousing	13.85	8.54	4.74	42.17	0.62
Information Media and Telecommunications	40.34	16.86	21.80	106.43	0.42
Financial and Insurance Services	29.75	13.67	15.04	93.41	0.46
Rental, Hiring and Real Estate Services	25.75	13.08	11.01	81.24	0.51
Professional, Scientific and Technical Services	19.90	6.26	13.10	48.09	0.31
Administrative and Support Services	1.91	0.88	1.05	5.75	0.46
Public Administration and Safety	74.79	33.14	30.94	222.61	0.44
Education and Training	74.60	32.48	45.57	225.26	0.44
Health Care and Social Assistance	39.38	16.76	24.67	118.76	0.43
Arts and Recreation Services	31.56	15.34	12.60	72.00	0.49
Other Services	9.59	5.63	5.55	34.46	0.59

Note: The table shows summary statistics for each of the series shown in Figure A B.2.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Figure A B.3. Ratio of the Lightcast to officially reported vacancies by unaggregated sector, quarterly aggregations, Canada

Note: Data for Q2 and Q3 in 2020 are missing due to the COVID-19 pandemic.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.
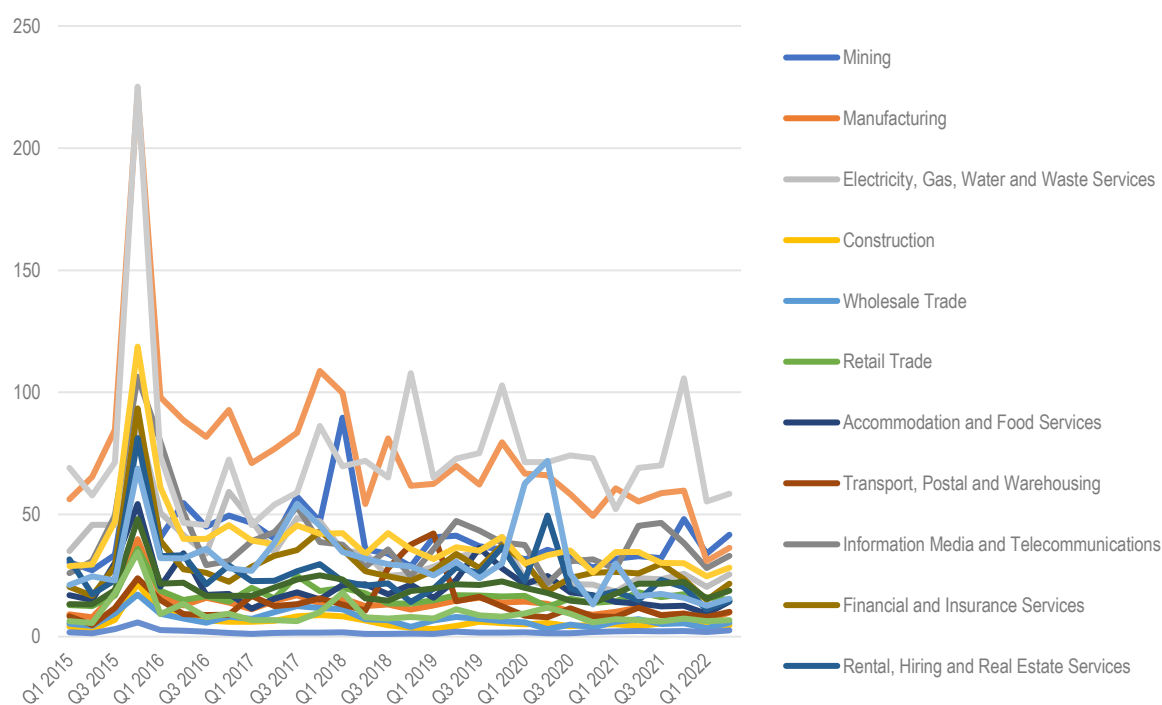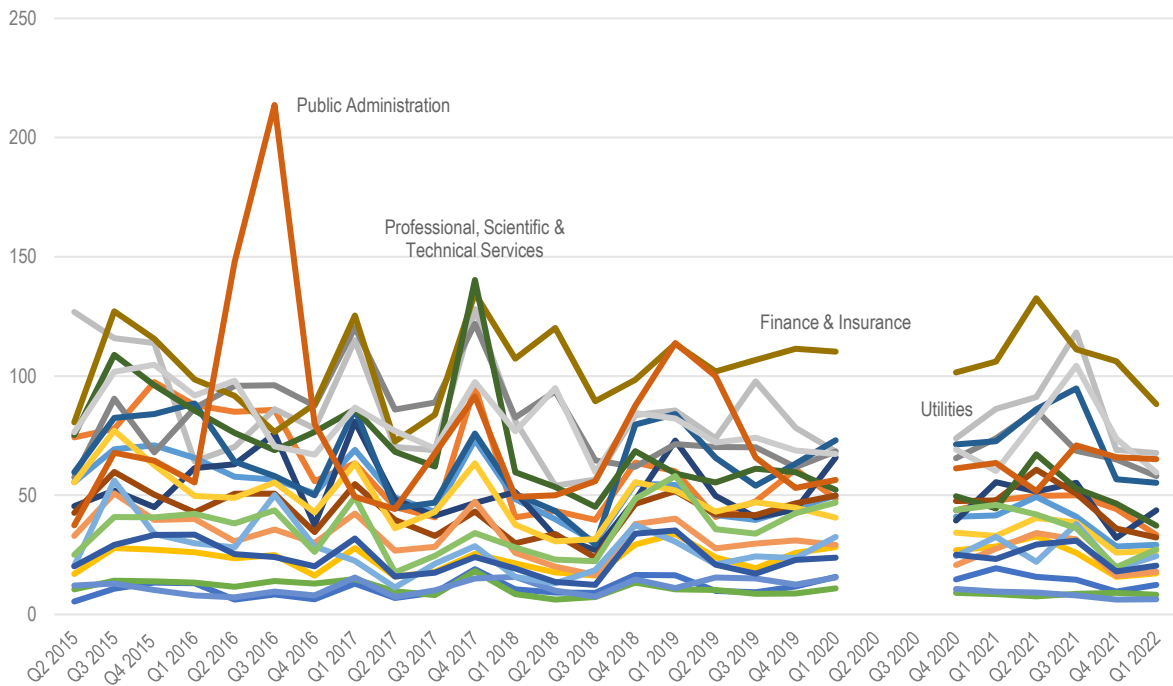
Table A B.3. Summary statistics for the ratio of vacancies by unaggregated sector, quarterly aggregations, Canada

Sector	Mean	Std. dev.	Min	Max	Coef. of var.
Agriculture, Forestry, Fishing and Hunting	11.78	3.92	5.38	19.35	0.33
Mining, Quarrying & Oil & Gas Extraction	58.54	19.13	33.33	97.45	0.33
Utilities	85.47	21.62	53.99	128.00	0.25
Construction	23.56	5.46	15.80	33.68	0.23
Manufacturing	49.78	12.66	28.44	73.19	0.25
Wholesale Trade	10.62	2.88	6.24	18.13	0.27
Retail Trade	50.23	13.37	27.01	80.88	0.27
Transportation & Warehousing	43.93	9.18	23.81	60.75	0.21
Information & Cultural Industries	79.22	17.26	57.47	121.85	0.22
Finance & Insurance	103.80	16.80	72.45	134.61	0.16
Real Estate & Rental & Leasing	66.11	16.86	29.30	94.79	0.26
Professional, Scientific & Technical Services	67.52	22.36	37.27	140.32	0.33
Management of Companies & Enterprises	10.89	3.21	6.28	15.69	0.29
Administrative & Support, Waste Management & Remediation Services	31.16	8.98	16.09	50.82	0.29
Educational Services	79.62	14.01	59.51	104.74	0.18
Health Care & Social Assistance	45.35	12.64	25.96	77.12	0.28
Arts, Entertainment & Recreation	27.21	10.26	11.26	56.07	0.38
Accommodation & Food Services	36.01	10.51	17.65	58.05	0.29
Other Services (except Public Administration)	23.87	6.57	12.35	35.20	0.28
Public Administration	74.17	37.06	37.39	213.61	0.50

Note: The table shows summary statistics and coefficient of variation for each of the series shown in Figure A B.3.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Figure A B.4. Ratio of the Lightcast to officially reported vacancies by unaggregated sector, quarterly aggregations, the UK

Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series) data.

Table A B.4. Summary statistics for the ratio of vacancies by unaggregated sector, quarterly aggregations, the UK

Sector	Mean	Std. dev.	Min	Max	Coef. of var.
Accommodation & Food Service	115.89	19.67	68.94	162.16	0.17
Administrative & Support Service	104.02	18.19	78.94	161.35	0.17
Arts, Entertainment & Recreation	144.78	38.38	79.55	206.30	0.27
Construction	103.78	16.57	70.04	142.39	0.16
Education	36.33	7.19	21.14	53.33	0.20
Electricity, Gas, Steam & Air Conditioning	135.37	32.50	88.97	225.61	0.24
Financial & Insurance	69.41	11.82	49.23	101.01	0.17
Human Health & Social Work	56.06	7.23	46.20	73.57	0.13
Information & Communication	110.64	20.92	71.14	160.55	0.19
Manufacturing	43.24	5.63	32.59	56.73	0.13
Mining & Quarrying	33.81	13.79	0.00	80.19	0.41
Other Service	93.90	21.72	49.10	134.61	0.23
Professional Scientific & Technical Activities	60.40	8.01	47.89	80.92	0.13
Public Admin & Defence; Compulsory Social Security	55.05	10.59	33.72	73.06	0.19
Real Estate	46.79	10.56	29.57	69.62	0.23
Transport & Storage	97.94	18.70	71.58	142.46	0.19
Water Supply, Sewerage, Waste & Remediation	62.49	14.91	41.55	111.88	0.24
Wholesale & Retail Trade; Repair of Motor Vehicles and Motor Cycles	137.29	25.22	96.57	197.02	0.18

Note: The table shows summary statistics for each of the series shown in Figure A B.4.
Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series) data.

Figure A B.5. Ratio of the Lightcast to officially reported vacancies by unaggregated sector, quarterly aggregations, the US

Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.
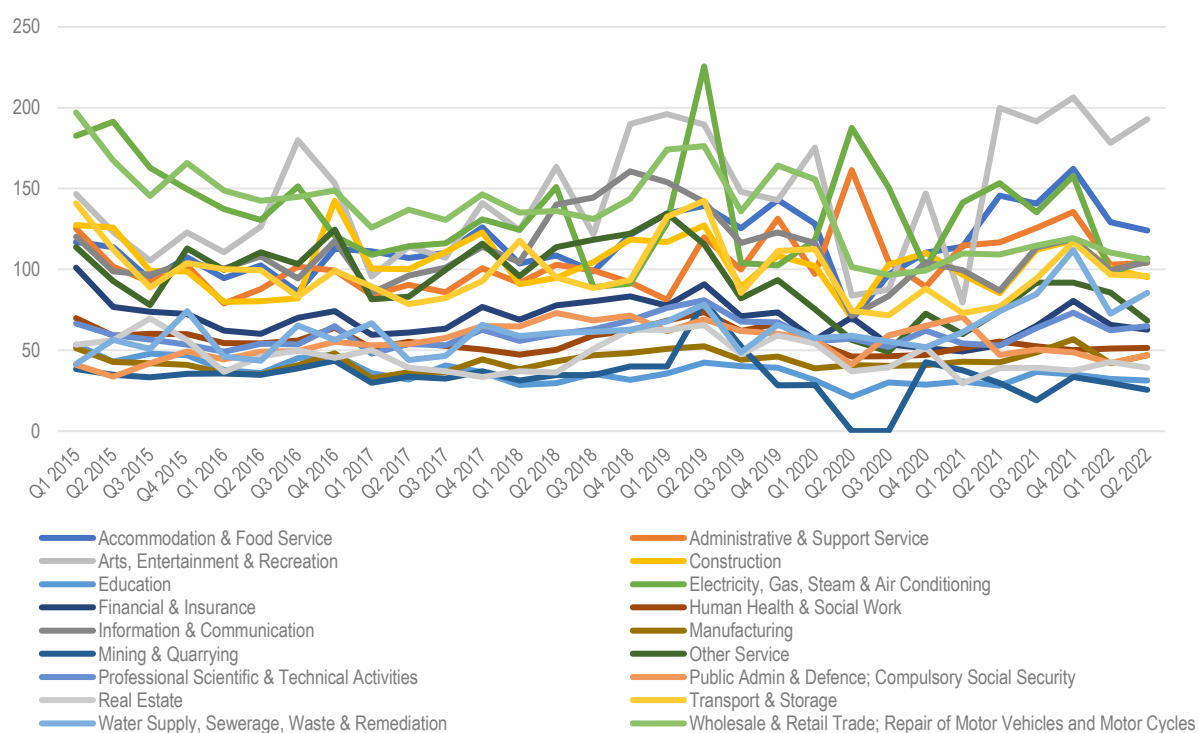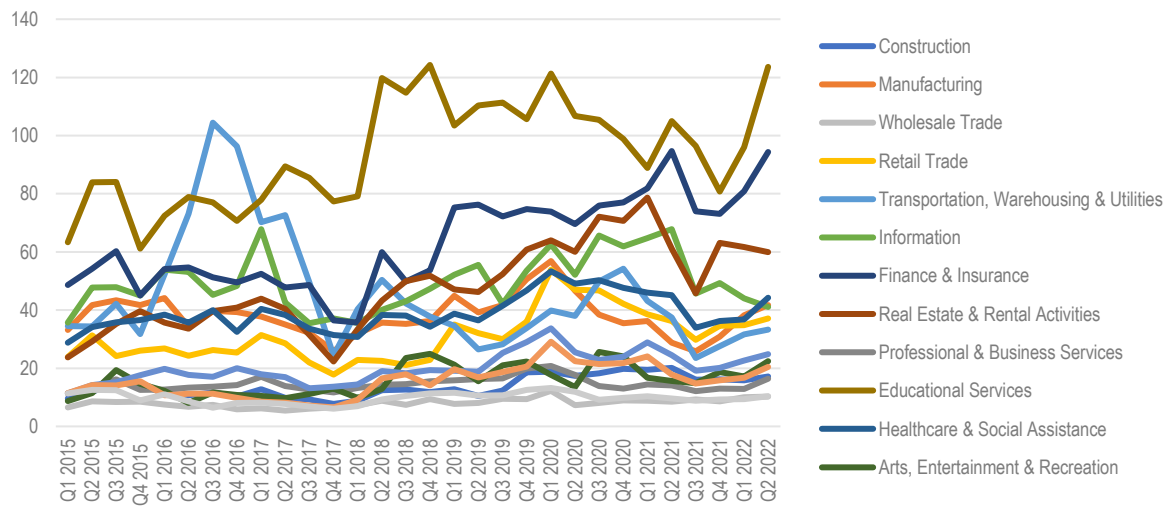
Table A B.5. Summary statistics for the ratio of vacancies by unaggregated sector, quarterly aggregations, the US

Sector	Mean	Std. dev.	Min	Max	Coef. of var.
Construction	13.81	3.67	7.74	20.14	0.27
Manufacturing	38.12	6.79	25.86	56.90	0.18
Wholesale Trade	8.15	1.48	5.49	12.21	0.18
Retail Trade	31.08	8.60	17.83	53.96	0.28
Transportation, Warehousing & Utilities	45.29	19.96	23.57	104.47	0.44
Information	49.49	9.68	35.50	67.80	0.20
Finance & Insurance	63.20	15.87	35.67	94.68	0.25
Real Estate & Rental Activities	47.94	14.70	22.35	78.71	0.31
Professional & Business Services	14.59	2.27	11.40	20.84	0.16
Educational Services	93.78	18.26	61.15	124.25	0.19
Healthcare & Social Assistance	39.13	6.19	28.84	53.34	0.16
Arts, Entertainment & Recreation	16.00	5.28	8.04	25.63	0.33
Accommodation & Food Services	20.16	5.15	11.42	33.70	0.26
Other Services	15.59	5.52	6.59	29.13	0.35
Government	9.82	1.91	6.14	13.20	0.19

Note: The table shows summary statistics and coefficient of variation for each of the series shown in Figure A B.5.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

Figure A B.6. Ratio of the Lightcast to officially-reported vacancies by region, quarterly aggregations, Australia

Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Table A B.6. Summary statistics for the ratio of vacancies by region, quarterly aggregations, Australia

State	Mean	Std. dev.	Min	Max
Australian Capital Territory	74.60	20.50	46.15	153.98
New South Wales	49.31	19.97	30.16	144.08
Northern Territory	45.48	21.41	23.29	146.70
Queensland	45.36	19.64	24.88	129.58
South Australia	40.50	14.16	20.50	89.13
Tasmania	47.03	19.62	28.61	137.70
Victoria	43.80	19.27	29.99	138.05
Western Australia	32.09	14.94	16.42	102.48

Note: The table shows summary statistics for each of the series shown in Figure A B.6.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Figure A B.7. Ratio of the Lightcast to officially-reported vacancies by region, quarterly aggregations, Canada

Note: Data for Q2 and Q3 in 2020 are missing due to the COVID-19 pandemic.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data

Table A B.7. Summary statistics for the ratio of vacancies by region, quarterly aggregations, Canada

State	Mean	Std. dev.	Min	Max
Alberta	75.24	16.94	45.45	113.84
British Columbia	62.59	13.50	37.27	93.99
Manitoba	67.45	18.29	36.02	98.48
New Brunswick	74.96	22.74	35.77	138.20
Newfoundland and Labrador	79.99	20.40	44.28	134.62
Northwest Territories	60.19	23.25	26.79	128.51
Nova Scotia	92.77	30.12	53.35	188.73
Nunavut	82.57	22.99	45.20	146.00
Ontario	69.29	13.76	41.26	96.03
Prince Edward Island	88.20	25.31	40.85	140.67
Quebec	59.85	29.08	17.99	119.25
Saskatchewan	149.83	36.23	64.01	201.59
Yukon	60.79	30.09	26.59	135.45

Note: The table shows summary statistics and coefficient of variation for each of the series shown in Figure A B.7.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Figure A B.8. Ratio of the Lightcast to officially-reported vacancies by region, quarterly aggregations, the US
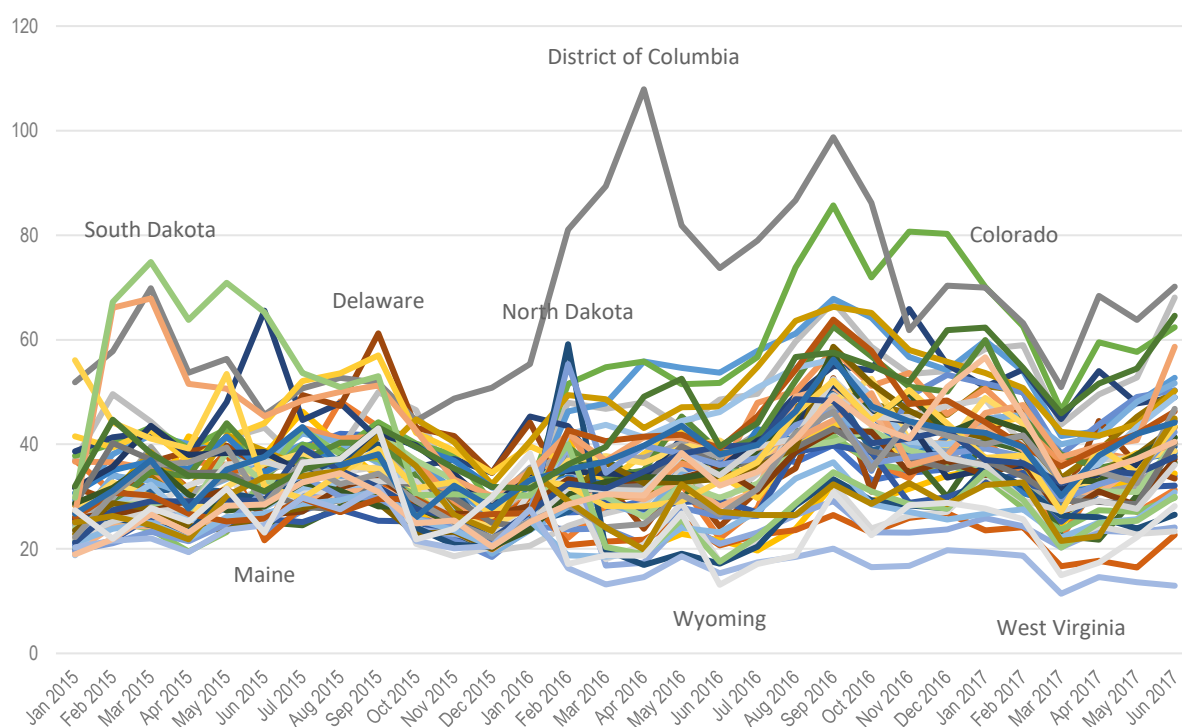
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.
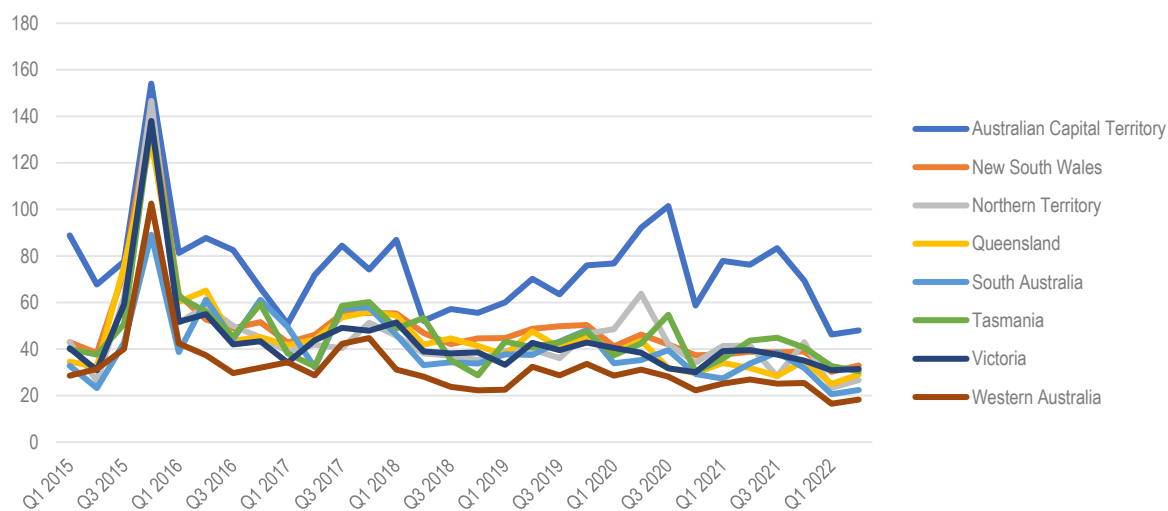
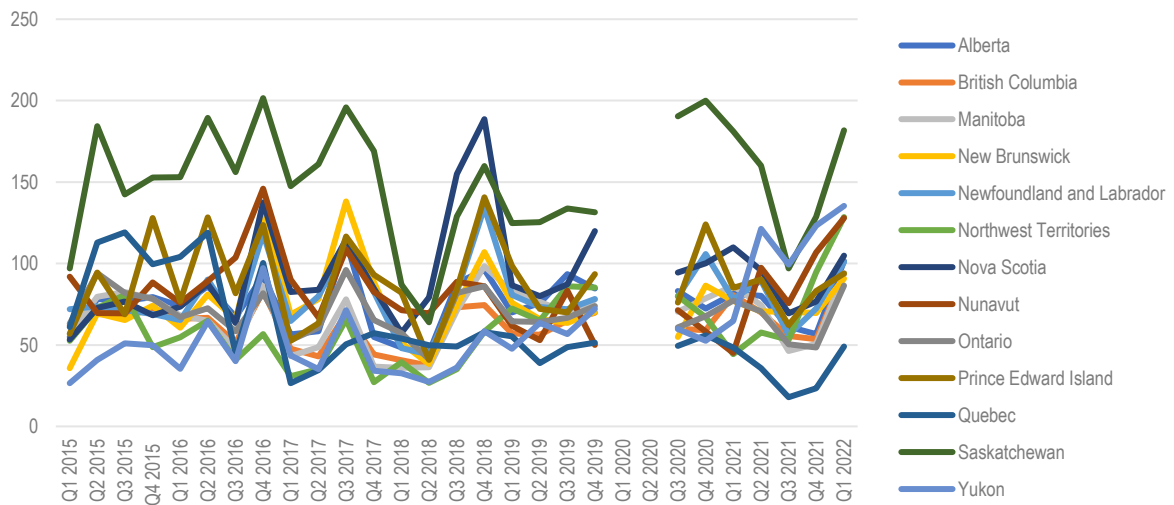## Table A B.8. Summary statistics for the ratio of vacancies by region, quarterly aggregations, the US

State	Mean	Std. dev.	Min	Max
Alabama	33.52	5.49	20.74	42.02
Alaska	33.14	6.97	21.98	48.17
Arizona	47.44	10.03	25.97	68.08
Arkansas	30.35	7.37	18.75	46.24
California	46.00	11.26	26.43	67.86
Colorado	52.40	15.54	28.86	85.73
Connecticut	45.61	8.69	33.40	65.89
Delaware	38.33	8.69	23.85	61.25
District of Columbia	66.44	16.49	44.55	107.92
Florida	35.67	9.43	20.67	51.15
Georgia	33.95	6.18	22.92	50.60
Hawaii	30.67	7.17	18.86	42.75
Idaho	33.25	10.94	18.42	53.25
Illinois	39.51	7.99	24.40	58.13
Indiana	33.83	8.28	19.59	49.40
Iowa	40.30	6.09	29.71	56.99
Kansas	36.11	7.92	19.06	51.65
Kentucky	30.85	6.25	19.49	42.95
Louisiana	28.21	5.17	19.66	39.84
Maine	25.41	5.37	16.42	38.45
Maryland	35.92	5.34	24.45	47.16
Massachusetts	45.30	10.12	28.29	66.34

Michigan	36.84	5.03	28.63	51.81
Minnesota	39.51	10.36	20.10	62.52
Mississippi	25.17	4.48	16.78	39.58
Missouri	32.23	8.35	19.34	50.97
Montana	26.71	5.78	17.54	42.40
Nebraska	36.37	6.91	24.11	52.52
Nevada	37.22	9.81	18.77	56.28
New Hampshire	33.21	5.16	23.87	42.76
New Jersey	37.71	5.04	27.88	51.52
New Mexico	31.89	4.73	25.86	42.63
New York	33.46	4.88	24.07	43.02
North Carolina	33.46	9.19	19.97	58.65
North Dakota	27.69	8.30	16.94	59.13
Ohio	34.93	8.25	20.34	57.45
Oklahoma	33.08	8.41	19.69	55.37
Oregon	43.96	9.33	26.64	67.91
Pennsylvania	33.10	7.04	21.90	52.48
Rhode Island	37.90	8.21	26.27	56.07
South Carolina	25.63	4.19	18.52	36.57
South Dakota	37.00	16.87	17.51	74.85
Tennessee	34.14	6.79	21.23	48.64
Texas	37.67	10.50	25.22	63.84
Utah	33.41	7.13	22.07	46.79
Vermont	29.61	5.89	19.85	44.95
Virginia	37.93	6.85	26.05	56.24
Washington	44.85	10.41	30.99	64.60
West Virginia	19.59	5.02	11.40	31.84
Wisconsin	33.62	9.33	18.92	56.64
Wyoming	25.48	7.24	13.13	43.07

Note: The table shows summary statistics and coefficient of variation for each of the series shown in Figure A B.8.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data

Annex C. Cross-walks from officially reported sectors in vacancy data to ten sectors used in the report

Table A C.1. Australia

Officially reported sectors	Aggregation into ten sectors
Accommodation and Food Services	Accommodation, Food, Arts & Recreation
Arts and Recreation Services	Accommodation, Food, Arts & Recreation
Construction	Construction
Education and Training	Education & Health
Health Care and Social Assistance	Education & Health
Financial and Insurance Services	FIRE
Rental, Hiring and Real Estate Services	FIRE
Information Media and Telecommunications	Information Media & Telecommunications
Electricity, Gas, Water and Waste Services	Manufacturing & Utilities
Manufacturing	Manufacturing & Utilities
Mining	Natural Resources
Administrative and Support Services	Professional, Scientific, Technical, Administrative & Other Services
Other Services	Professional, Scientific, Technical, Administrative & Other Services
Professional, Scientific and Technical Services	Professional, Scientific, Technical, Administrative & Other Services
Public Administration and Safety	Public Administration and Safety
Retail Trade	Trades, Transport & Warehousing
Transport, Postal and Warehousing	Trades, Transport & Warehousing
Wholesale Trade	Trades, Transport & Warehousing

Source: OECD elaboration.

Table A C.2. Canada

Officially reported sectors	Aggregation into ten sectors
Accommodation & Food Services	Accommodation, Food, Arts & Recreation
Arts, Entertainment & Recreation	Accommodation, Food, Arts & Recreation
Construction	Construction
Educational Services	Education & Health
Health Care & Social Assistance	Education & Health
Finance & Insurance	FIRE
Real Estate & Rental & Leasing	FIRE
Information & Cultural Industries	Information & Communication
Manufacturing	Manufacturing & Utilities
Utilities	Manufacturing & Utilities
Agriculture, Forestry, Fishing & Hunting	Natural Resources & Agriculture
Mining, Quarrying & Oil and Gas Extraction	Natural Resources & Agriculture
Administrative & Support, Waste Management & Remediation	Professional, Scientific, Technical, Administrative & Other Services
Management of Companies and Enterprises	Professional, Scientific, Technical, Administrative & Other Services
Other Services (except Public Administration)	Professional, Scientific, Technical, Administrative & Other Services
Professional, Scientific & Technical Services	Professional, Scientific, Technical, Administrative & Other Services
Public Administration	Public Administration
Retail Trade	Trades, Transport & Equipment Operators
Transportation & Warehousing	Trades, Transport & Equipment Operators
Wholesale Trade	Trades, Transport & Equipment Operators

Source: OECD elaboration.

Table A C.3. The UK

Officially reported sectors	Aggregation into ten sectors
Accommodation & Food	Accommodation, Food, Arts & Recreation
Arts, Entertainment & Recreation	Accommodation, Food, Arts & Recreation
Construction	Construction
Education	Education & Health
Human Health & Social Work	Education & Health
Financial & Insurance	FIRE
Real Estate	FIRE
Information & Communication	Information & Communication
Electricity, Gas, Steam & Air Conditioning	Manufacturing & Utilities
Manufacturing	Manufacturing & Utilities
Water, Sewerage, Waste & Remediation	Manufacturing & Utilities
Mining & Quarrying	Natural Resources
Administrative & Support Services	Professional, Scientific, Technical, Administrative & Other Services
Other Services	Professional, Scientific, Technical, Administrative & Other Services
Professional Scientific & Technical	Professional, Scientific, Technical, Administrative & Other Services
Public Admin & Defence; Compulsory Social Security	Public Administration & Safety
Transport & Storage	Trades, Transport & Equipment Operators
Wholesale & Retail; Repair of Motor Vehicles and Motorcycles	Trades, Transport & Equipment Operators

Source: OECD elaboration.

Table A C.4. The US

Officially reported sectors	Aggregation into ten sectors
Accommodation and food services	Accommodation, Food, Arts & Recreation
Arts, entertainment and recreation	Accommodation, Food, Arts & Recreation
Construction	Construction
Educational services	Education, Healthcare & Social Assistance
Healthcare and social assistance	Education, Healthcare & Social Assistance
Finance and insurance	FIRE
Real estate and rental activities	FIRE
Government	Government
Information	Information
Manufacturing	Manufacturing
Other services	Professional, Scientific, Technical & Other Services
Professional, scientific and technical services	Professional, Scientific, Technical & Other Services
Transportation, warehousing and utilities	Transportation, Warehousing & Utilitites
Retail trade	Wholesale & Retail Trade
Wholesale trade	Wholesale & Retail Trade

Source: OECD elaboration.

Annex D. Comparions of distributions across sectors in 2022, no aggregations

Figure A D.1. Comparison of distributions across sectors in 2022 (no aggregation), Australia

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
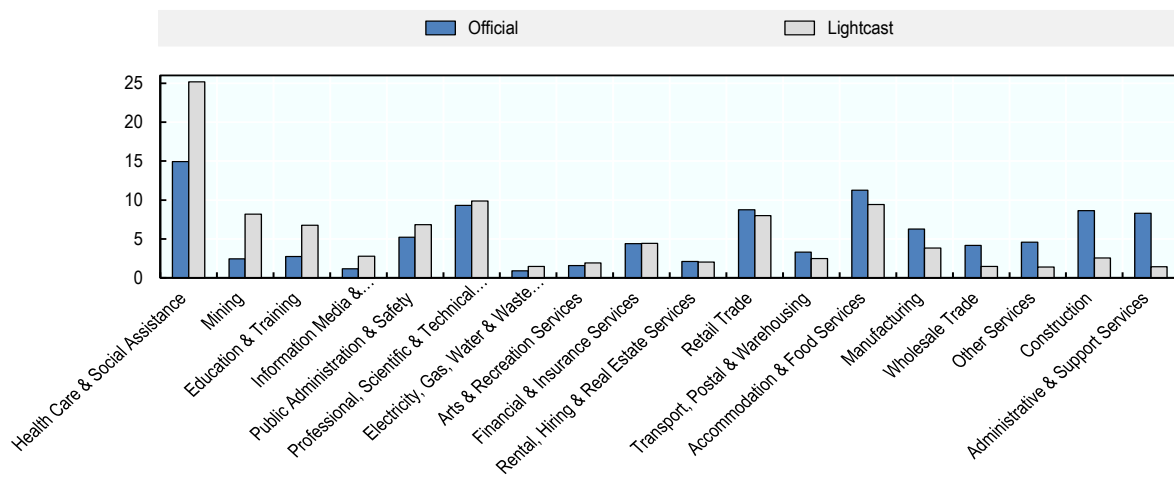
Note: Data for 21 April – 20 May 2022; sectors are ordered from the most overrepresented to the most underrepresented in the Lightcast data.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly, 2022) data.

Figure A D.2. Comparison of distributions across sectors in 2022 (no aggregation), Canada

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.

Note: Data for Q1 2022; sectors are ordered from the most overrepresented in the Lightcast data to the most underrepresented.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey, 2022) data.

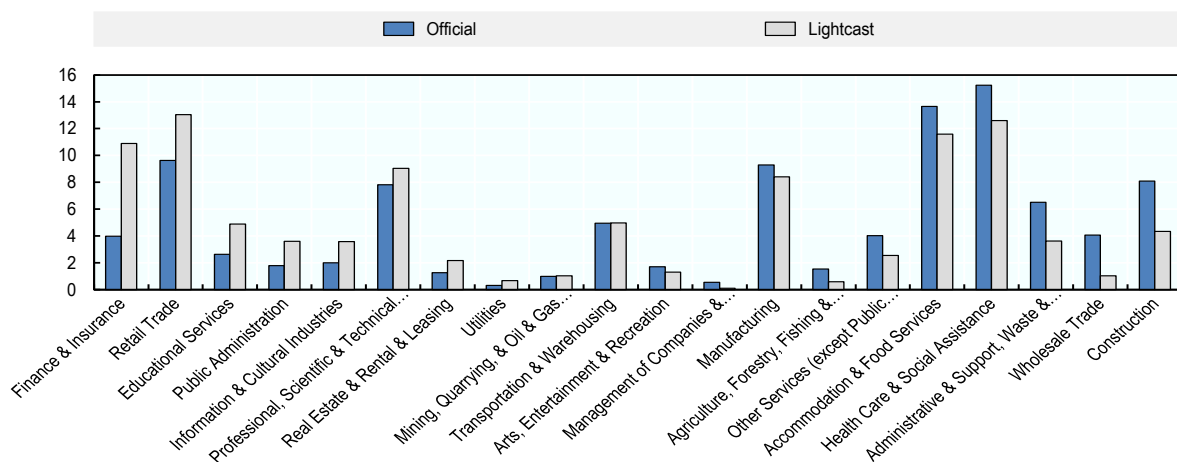### Figure A D.3. Comparison of distributions across sectors in 2022 (no aggregation), the UK

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.

Note: Data for Q2 2022; sectors are ordered from the most overrepresented in the Lightcast data to the most underrepresented.
Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series, 2022) data.

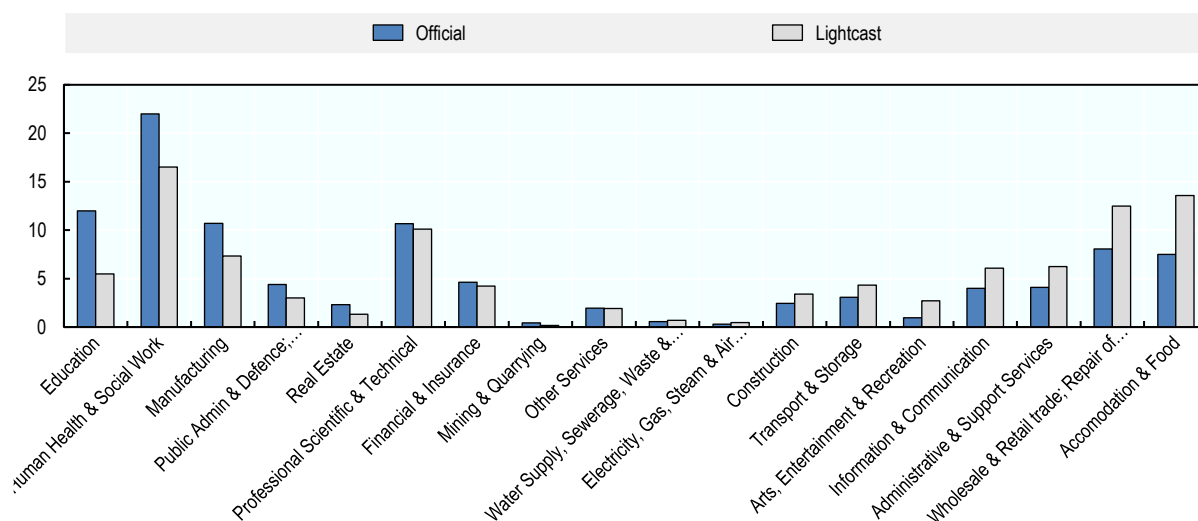### Figure A D.4. Comparison of distributions across sectors in 2022 (no aggregation), the US

Share of officially reported vacancies (Official) vs. share of Lightcast vacancies, %.
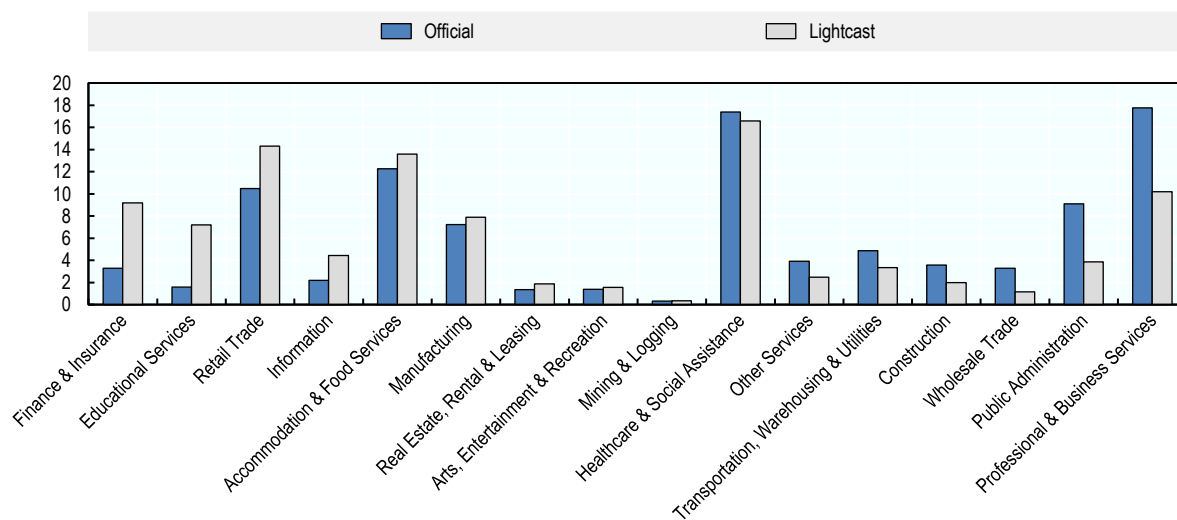
Note: Data for May 2022; sectors are ordered from the most overrepresented in the Lightcast data to the most underrepresented.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey, 2022) data.

Annex E. Comparisons of sectoral distributions over time, no aggregations

Figure A E.1. Representativeness by unaggregated industrial sectors over time, Australia

Differences between vacancy shares reported by the Lightcast data and the official data.
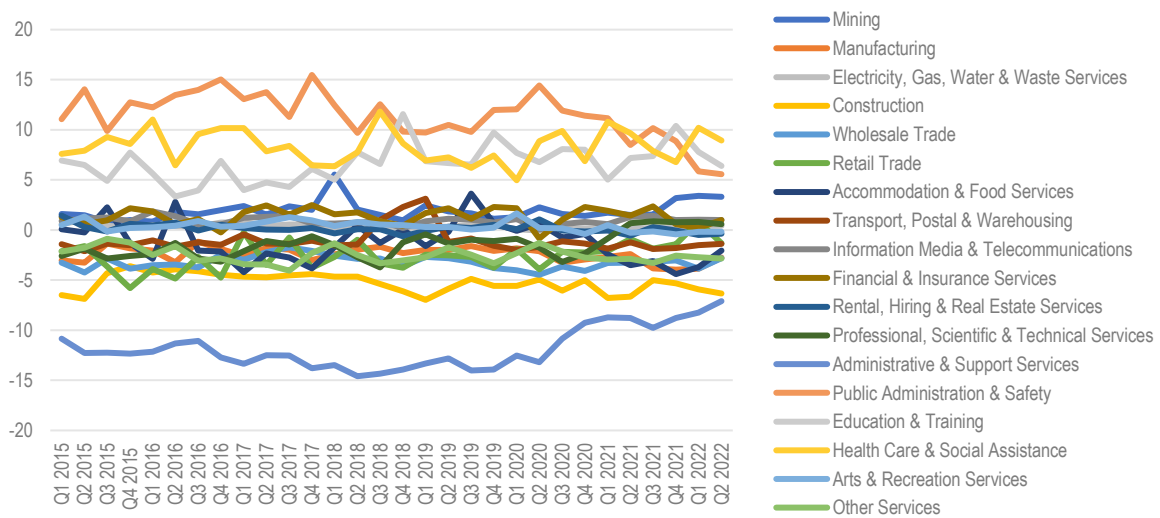
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Table A E.1. Summary statistics for differences in shares, unaggregated sectors, Australia

| Sector	Mean	Std. dev.	Min	Max
Mining	1.92	0.96	0.77	5.53
Manufacturing	-2.43	0.75	-3.95	-1.17
Electricity, Gas, Water & Waste Services	0.36	0.24	-0.04	1.03
Construction	-5.27	0.95	-6.98	-3.62
Wholesale Trade	-3.21	0.66	-4.47	-1.72
Retail Trade	-2.56	1.46	-5.81	0.85
Accommodation & Food Services	-1.21	2.04	-4.41	3.62
Transport, Postal & Warehousing	-1.07	1.17	-2.10	3.12
Information, Media & Communications	0.94	0.38	0.03	1.84
Financial & Insurance Services	1.35	0.84	-0.84	2.50
Rental, Hiring and Real Estate Services	0.14	0.46	-0.63	1.44
Professional, Scientific & Technical Services	-1.44	1.29	-3.74	0.87
Administrative & Support Services	-11.83	2.05	-14.60	-7.10
Public Administration & Safety	11.41	2.39	5.56	15.48
Education & Training	6.68	1.87	3.34	11.56
Health Care & Social Assistance	8.36	1.66	4.97	11.81
Arts & Recreation Services	0.38	0.51	-0.42	1.65
Other Services	-2.51	0.76	-4.05	-0.88

Note: The table shows summary statistics for each of the series shown in Figure A E.1.
Source: OECD calculations based on the Lightcast data and the Australian Bureau of Statistics (Labour Force Quarterly) data.

Figure A E.2. Representativeness by unaggregated industrial sectors over time, Canada

Differences between vacancy shares reported by the Lightcast data and the official data.
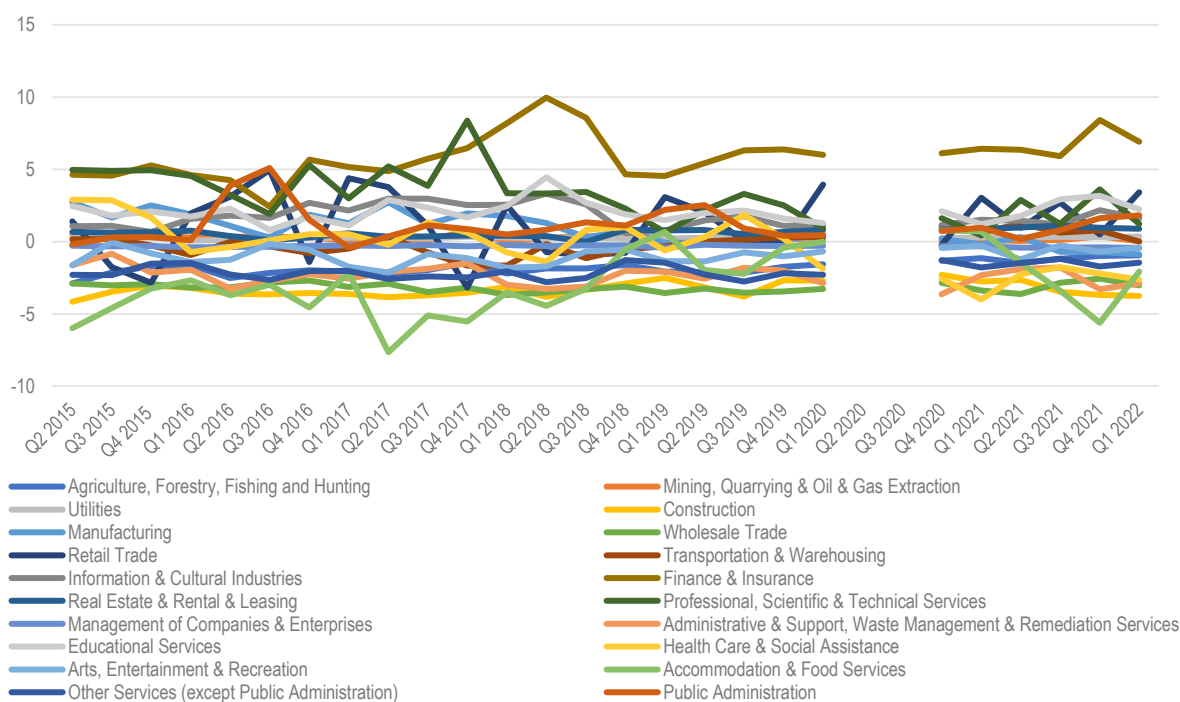
Note: Data for Q2 and Q3 in 2020 are missing due to the COVID-19 pandemic.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Table A E.2. Summary statistics for differences in shares, unaggregated sectors, Canada
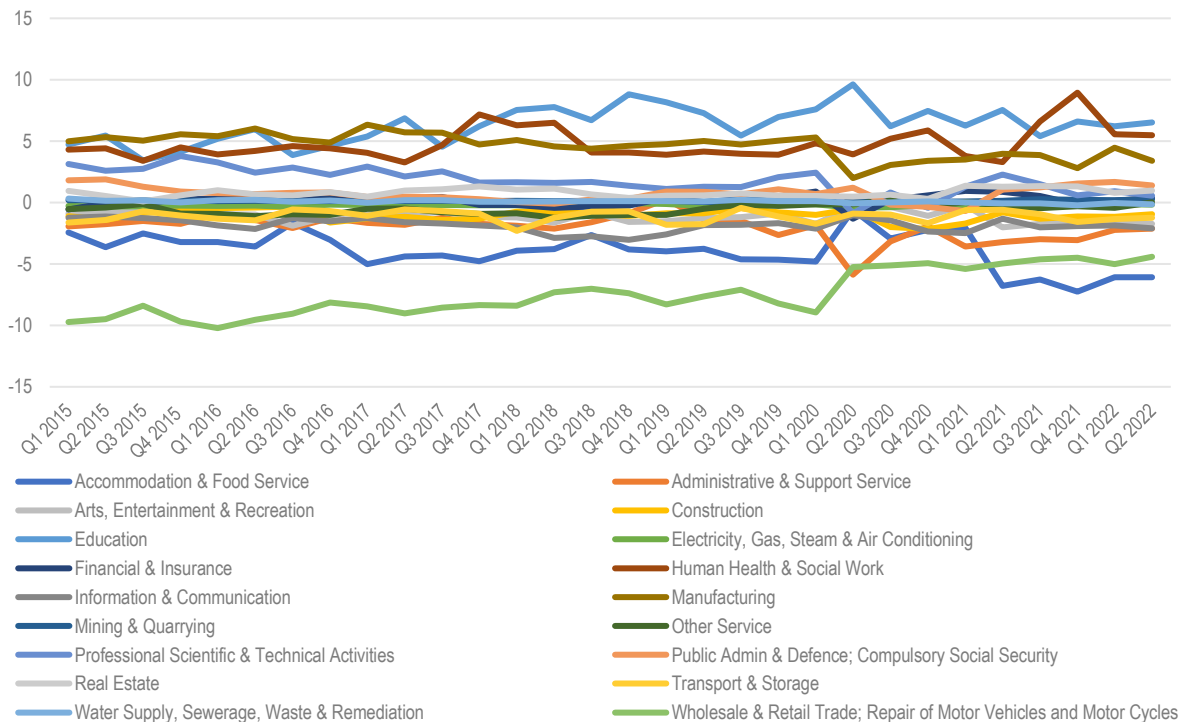
Sector	Mean	Std. dev.	Min	Max
Agriculture, Forestry, Fishing and Hunting	-1.81	0.47	-2.89	-0.95
Mining, Quarrying & Oil & Gas Extraction	0.21	0.18	-0.03	0.73
Utilities	0.27	0.09	0.08	0.54
Construction	-3.31	0.49	-4.15	-2.30
Manufacturing	0.83	1.08	-0.88	2.74
Wholesale Trade	-3.17	0.30	-3.70	-2.56
Retail Trade	1.19	2.26	-3.18	4.97
Transportation & Warehousing	-0.12	0.74	-1.66	1.24
Information & Cultural Industries	1.75	0.77	0.58	3.32
Finance & Insurance	5.92	1.59	2.42	9.96
Real Estate & Rental & Leasing	0.62	0.28	-0.01	1.11
Professional, Scientific & Technical Services	3.20	1.80	0.43	8.39
Management of Companies & Enterprises	-0.32	0.06	-0.43	-0.22
Administrative & Support, Waste Management & Remediation Services	-2.36	0.67	-3.64	-0.84
Educational Services	2.09	0.75	0.81	4.46
Health Care & Social Assistance	-0.25	1.74	-3.99	2.90
Arts, Entertainment & Recreation	-0.96	0.57	-2.15	-0.04
Accommodation & Food Services	-2.88	2.24	-7.64	0.79
Other Services (except Public Administration)	-2.02	0.50	-2.80	-1.18
Public Administration	1.13	1.22	-0.40	5.10

Note: The table shows summary statistics for each of the series shown in Figure A E.2.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Figure A E.3. Representativeness by unaggregated industrial sectors over time, the UK

Differences between vacancy shares reported by the Lightcast data and the official data.

Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series) data.

Table A E.3. Summary statistics for differences in shares, unaggregated sectors, the UK
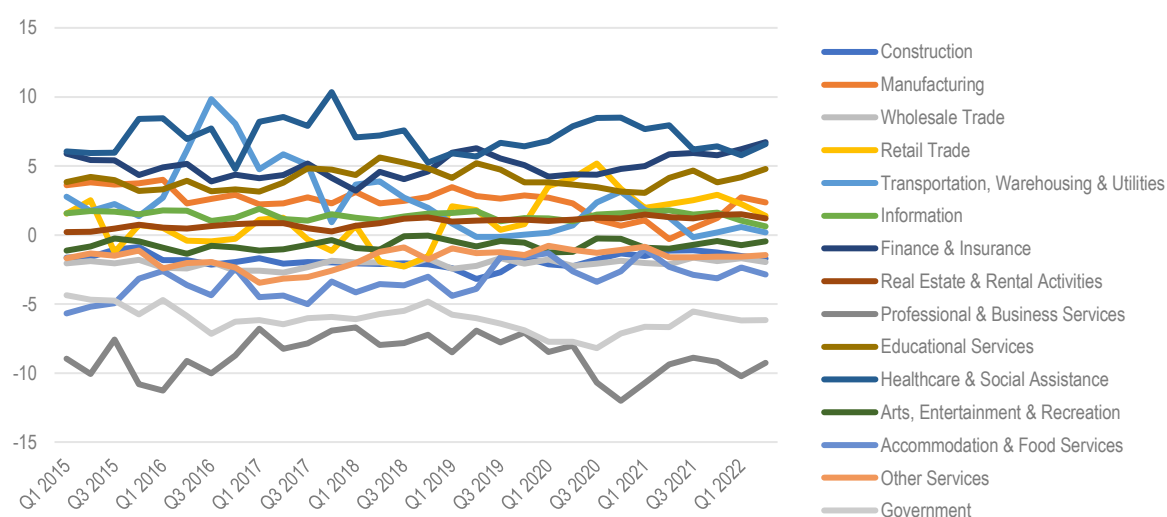
Sector	Mean	Std. dev.	Min	Max
Accommodation & Food Service	-3.94	1.55	-7.25	-0.59
Administrative & Support Service	-2.03	1.07	-5.88	0.28
Arts, Entertainment & Recreation	-1.14	0.49	-2.01	-0.22
Construction	-1.05	0.45	-2.15	-0.34
Education	6.28	1.47	3.43	9.63
Electricity, Gas, Steam & Air Conditioning	-0.23	0.11	-0.45	-0.05
Financial & Insurance	0.17	0.48	-1.32	0.94
Human Health & Social Work	4.77	1.29	3.27	8.93
Information & Communication	-1.85	0.53	-3.03	-1.04
Manufacturing	4.62	1.00	1.99	6.33
Mining & Quarrying	0.15	0.05	0.04	0.26
Other Service	-0.56	0.40	-1.24	0.25
Professional Scientific & Technical Activities	1.78	1.06	-0.89	3.78
Public Admin & Defence; Compulsory Social Security	0.74	0.63	-0.71	1.88
Real Estate	0.79	0.34	0.09	1.35
Transport & Storage	-1.13	0.47	-2.32	-0.42
Water Supply, Sewerage, Waste & Remediation	0.07	0.12	-0.22	0.32
Wholesale & Retail Trade; Repair of Motor Vehicles and Motor Cycles	-7.44	1.85	-10.22	-4.42

Note: The table shows summary statistics for each of the series shown in Figure A E.3.
Source: OECD calculations based on the Lightcast data and the Office for National Statistics (Labour Market Statistics Time Series) data.

Figure A E.4. Representativeness by unaggregated industrial sectors over time, the US

Differences between vacancy shares reported by the Lightcast data and the official data.

Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

Table A E.4. Summary statistics for differences in shares, unaggregated sectors, the US

Sector	Mean	Std. dev.	Min	Max
Construction	-1.82	0.49	-3.18	-0.88
Manufacturing	2.43	1.03	-0.31	4.00
Wholesale Trade	-2.08	0.29	-2.72	-1.64
Retail Trade	1.11	1.83	-2.28	5.18
Transportation, Warehousing & Utilities	2.48	2.52	-0.17	9.83
Information	1.40	0.32	0.62	1.90
Finance & Insurance	4.99	0.84	3.20	6.72
Real Estate & Rental Activities	0.91	0.38	0.21	1.49
Professional & Business Services	-8.77	1.45	-12.00	-6.69
Educational Services	4.07	0.71	3.04	5.60
Healthcare & Social Assistance	7.11	1.23	4.79	10.36
Arts, Entertainment & Recreation	-0.72	0.36	-1.37	-0.04
Accommodation & Food Services	-3.32	1.19	-5.68	-1.01
Other Services	-1.68	0.69	-3.46	-0.78
Government	-6.11	0.93	-8.18	-4.35

Note: The table shows summary statistics for each of the series shown in Figure A E.4.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

Annex F. Additional comparisons of distributions, monthly aggregations

Figure A F.1. Representativeness by ten sectors over time, monthly data, Canada

Differences between vacancy shares reported by the Lightcast data and the official data.

Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Table A F.1. Summary statistics for differences in shares of ten sectors, monthly data, Canada

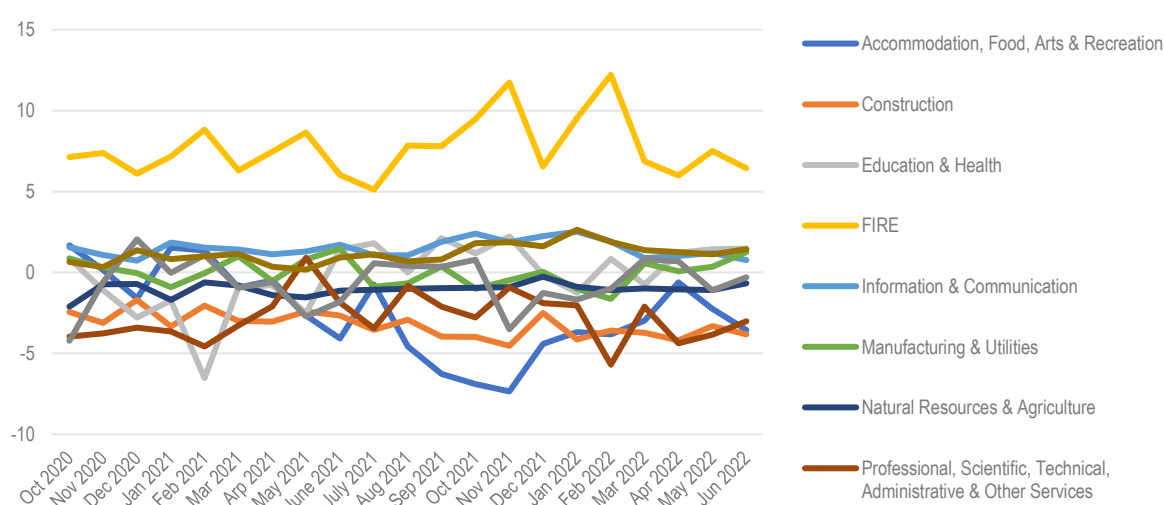| Sector	Mean	Std. dev.	Min	Max
Accommodation, Food, Arts & Recreation	-2.49	2.66	-7.34	1.67
Construction	-3.24	0.76	-4.54	-1.69
Education & Health	-0.18	2.09	-6.53	2.23
FIRE	7.72	1.82	5.12	12.21
Information & Communication	1.48	0.53	0.72	2.53
Manufacturing & Utilities	-0.01	0.84	-1.64	1.48
Natural Resources & Agriculture	-1.03	0.40	-2.12	-0.26
Professional, Scientific, Technical, Administrative & Other Services	-2.80	1.48	-5.69	0.91
Trades, Transport, Warehousing & Equipment Operators	-0.62	1.56	-4.21	2.04
Public Administration	1.16	0.60	0.17	2.63

Note: The table shows summary statistics for each of the series shown in Figure A F.1.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Figure A F.2. Representativeness by ten sectors over time, monthly data, the US

Differences between vacancy shares reported by the Lightcast data and the official data.

Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

Table A F.2. Summary statistics for differences in shares of ten sectors, monthly data, the US

Sectors	Mean	Std. dev.	Min	Max
Government	-6.12	1.11	-8.63	-3.48
Construction	-1.82	0.60	-3.76	-0.51
Manufacturing	2.43	1.10	-0.54	4.49
Wholesale & Retail Trade	-0.95	2.19	-5.20	4.64
Transportation, Warehousing & Utilities	2.44	2.71	-1.35	12.71
Information	1.40	0.41	0.41	2.50
FIRE	5.90	1.16	3.23	8.64
Professional, Scientific, Technical & Other Services	-10.44	1.72	-14.93	-5.95
Education, Healthcare & Social Assistance	11.19	1.78	6.67	17.29
Accommodation, Food, Arts & Recreation	-4.04	1.62	-8.05	-0.26

Note: The table shows summary statistics for each of the series shown in Figure A F.2.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

Figure A F.3. Representativeness by region, monthly data, Canada

Differences between vacancy shares reported by the Lightcast data and the official data.

Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

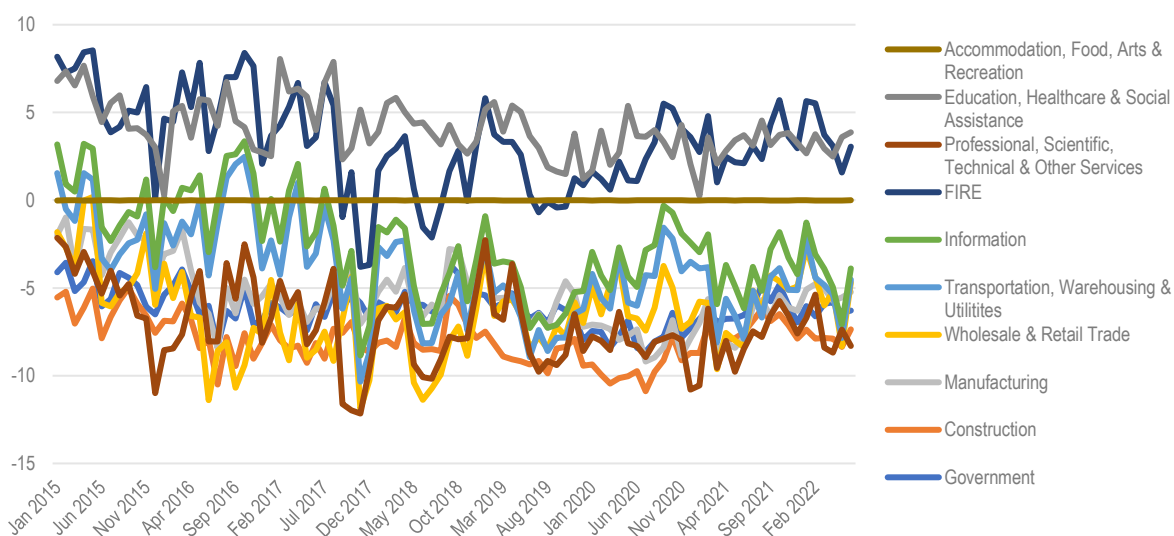## Table A F.3. Summary statistics for differences in shares by region, monthly data, Canada

Region	Mean	Std. dev.	Min	Max
Alberta	1.87	1.64	-1.25	6.25
British Columbia	1.03	2.07	-4.43	4.52
Manitoba	0.46	0.36	-0.18	1.19
New Brunswick	0.37	0.45	-0.42	1.25
Newfoundland and Labrador	0.26	0.16	-0.08	0.58
Northwest Territories	0.03	0.08	-0.06	0.26
Nova Scotia	1.09	0.29	0.41	1.51
Nunavut	0.01	0.05	-0.07	0.08
Ontario	1.50	2.58	-3.85	6.93
Prince Edward Island	0.15	0.11	-0.08	0.43
Quebec	-10.19	4.55	-17.44	-0.05
Saskatchewan	3.28	0.90	1.26	4.73
Yukon	0.09	0.13	-0.07	0.35

Note: The table shows summary statistics for each of the series shown in Figure A F.3.
Source: OECD calculations based on the Lightcast data and the Statistics Canada (Job Vacancy and Wage Survey) data.

Figure A F.4. Representativeness by region, monthly data, the US

Differences between vacancy shares reported by the Lightcast data and the official data.

Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.

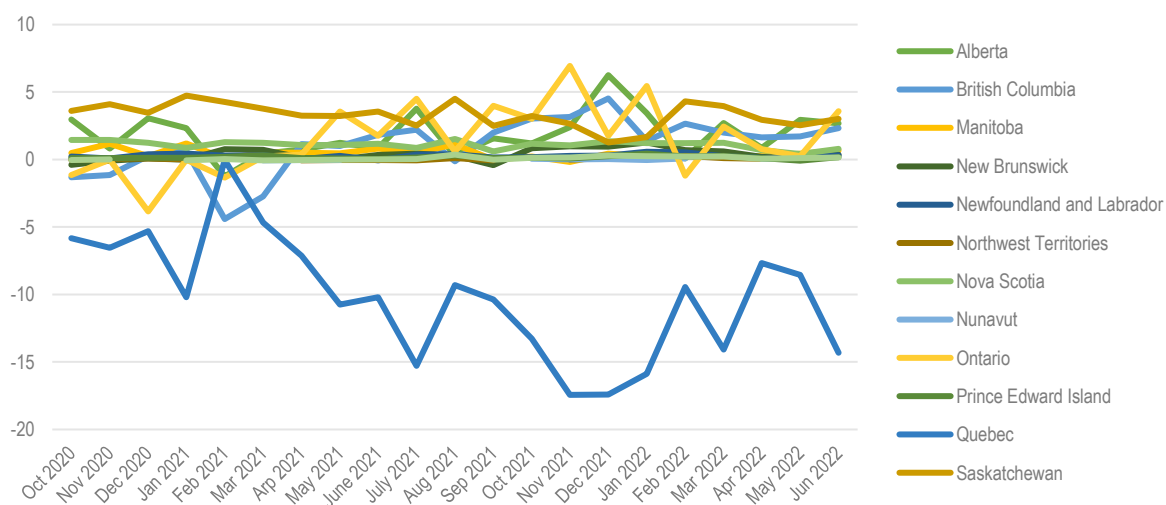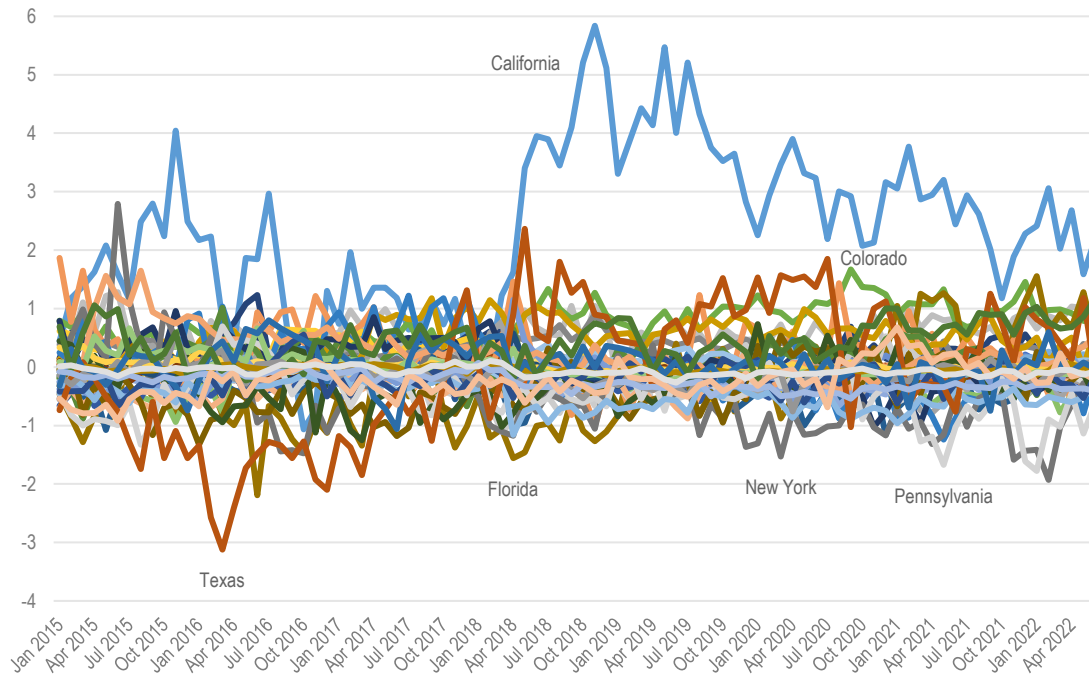### Table A F.4. Summary statistics for differences in shares by region, monthly data, the US

Region	Mean	Std. dev.	Min	Max
Alabama	-0.12	0.23	-0.65	0.59
Alaska	-0.03	0.09	-0.26	0.15
Arizona	0.57	0.31	-0.22	1.28
Arkansas	-0.14	0.22	-0.57	0.61
California	2.39	1.42	-1.06	5.84
Colorado	0.74	0.36	0.01	1.67
Connecticut	0.24	0.24	-0.18	1.23
Delaware	0.02	0.10	-0.20	0.26
District of Columbia	0.34	0.14	0.10	0.90
Florida	-0.31	0.77	-2.19	1.56
Georgia	-0.30	0.30	-1.16	0.35
Hawaii	-0.08	0.07	-0.32	0.08
Idaho	-0.06	0.12	-0.33	0.22
Illinois	0.23	0.54	-1.00	1.86
Indiana	-0.21	0.29	-1.10	0.53
Iowa	0.12	0.28	-0.43	0.65
Kansas	-0.04	0.16	-0.39	0.33
Kentucky	-0.24	0.25	-0.94	0.62
Louisiana	-0.32	0.14	-0.68	-0.02
Maine	-0.13	0.11	-0.35	0.14
Maryland	-0.05	0.28	-0.78	0.77
Massachusetts	0.51	0.32	-0.36	1.18

Michigan	0.02	0.54	-1.24	1.22
Minnesota	0.08	0.31	-0.88	1.02
Mississippi	-0.24	0.16	-0.66	0.45
Missouri	-0.29	0.22	-0.88	0.26
Montana	-0.11	0.09	-0.27	0.28
Nebraska	-0.02	0.08	-0.19	0.16
Nevada	-0.01	0.16	-0.44	0.40
New Hampshire	-0.05	0.08	-0.23	0.14
New Jersey	0.06	0.39	-1.02	0.96
New Mexico	-0.08	0.08	-0.25	0.10
New York	-0.50	0.72	-1.93	2.79
North Carolina	-0.38	0.41	-1.33	0.54
North Dakota	-0.08	0.10	-0.33	0.57
Ohio	-0.27	0.40	-1.25	0.74
Oklahoma	-0.14	0.22	-0.61	1.15
Oregon	0.26	0.40	-0.48	1.65
Pennsylvania	-0.46	0.46	-1.78	0.70
Rhode Island	0.01	0.09	-0.11	0.36
South Carolina	-0.49	0.23	-1.16	-0.03
South Dakota	0.01	0.19	-0.23	0.67
Tennessee	-0.18	0.23	-0.66	0.55
Texas	-0.05	1.17	-3.12	2.37
Utah	-0.11	0.18	-0.76	0.55
Vermont	-0.04	0.06	-0.20	0.09
Virginia	0.06	0.33	-1.06	0.80
Washington	0.43	0.35	-0.24	1.17
West Virginia	-0.26	0.13	-0.54	0.08
Wisconsin	-0.23	0.29	-0.92	0.68
Wyoming	-0.07	0.07	-0.26	0.11

Note: The table shows summary statistics for each of the series shown in Figure A F.4.
Source: OECD calculations based on the Lightcast data and the Bureau of Labour Statistics (Job Openings and Labor Turnover Survey) data.