



OECD Local Economic and Employment Development (LEED)
Papers 2024/02

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

**Wessel Vermeulen,
Fernanda Gutierrez
Amaros**

<https://dx.doi.org/10.1787/e1026d81-en>

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

By Wessel Vermeulen and Fernanda Gutierrez Amaros

Data on online job postings represents an important source of information for local labour markets. Many countries lack statistics on labour demand that are sufficiently up-to-date and disaggregated across regions, sectors and occupations. Web-scraped data from online job postings can provide further insights on the trends in labour demand and the skills needed across regions, sectors and occupations. This paper assesses the comparability and validity between Lightcast and other data sources for Austria, Belgium, Bulgaria, Germany, Hungary, the Netherlands, Portugal, Romania, Spain and Sweden, for the years 2019 to 2022 across regions, sectors and occupations. It concludes with some recommendations for labour market analysts that want to use data on online job postings for assessing labour demand trends.

JEL codes: C89; J23, J63, J29; O50; R12; Y1

Keywords: online job postings; vacancy data; Lightcast (Burning Glass); unconventional data sources; big data

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.

© OECD 2024

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <https://www.oecd.org/termsandconditions>.

Table of contents

Acknowledgements	5
Executive summary	6
1 Introduction	8
Workers increasingly use online tools in their job search	8
Online job searching provides a new data source for labour market analysis	10
2 Data and methodology	11
Lightcast online job postings data	11
Eurostat data	11
National sources	12
Job postings, open vacancies and surveys	13
Methodology	14
3 Results	15
The coverage of Lightcast data across countries and time	15
The coverage of Lightcast data across subnational regions	18
Sectoral coverage	20
Occupational coverage	22
Conditional correlations for multi-dimensional vacancy counts	24
Discussion and conclusion	28
References	30
Annex A. National sources and matching data dimensions.	32

FIGURES

Figure 1. Using internet for job applications is increasing across most EU countries	9
Figure 2. Lightcast online job postings rise rapidly between 2018 and 2019 for many countries	15
Figure 3. Lightcast and Eurostat do not have equal distributions across countries	16
Figure 4. Interpretation of the bar graphs	17
Figure 5. The number of online job postings relative to national sources varies across regions for most countries	19
Figure 6. Sectoral distributions of vacancies between online job postings and national sources	21
Figure 7. The ratio of online job postings over aggregate EU vacancies across sectors varies over time and sector	22

Figure 8. Occupational distributions of vacancies between Lightcast and national sources	23
--	----

TABLES

Table 1. Overview of vacancy data from national statistical sources used in this paper	12
Table 2. Three sources, three ways to count vacancies	13
Table 3. Sectors according to NACE classification	20
Table 4. ISCO occupations	22
Table 5. A national source vacancy equals between 0.12 to 9.8 vacancies in Lightcast	25
Table 6. At higher levels of aggregation of occupations, Lightcast and national sources correlate more	26
Table A A.1. Austrian occupation classification	32
Table A A.2 Occupation classification for Austria	32
Table A A.3. Portugal's sector classification	33

Acknowledgements

This paper was prepared within the framework of the Spatial Productivity Lab (SPL) at the OECD Trento Centre for Local Development. The Centre is a part of the Local Economic and Employment Development (LEED) Programme at the OECD Centre for Entrepreneurship, SMEs, Regions and Cities (CFE), led by Lamia Kamal-Chaoui, Director.

The authors thank Nadim Ahmad, Rudiger Ahrend, Karen Maguire, Kristine Langenbucher, Lars Ludolph (all CFE), Annabelle Mourougane (Statistics and Data Directorate), Fabio Manca (Directorate for Employment, Labour and Social Affairs), internal OECD seminar participants and Mauro Pelucchi (Lightcast) for helpful comments and suggestions. The authors are also grateful for the help of Elisa Campestrin (OECD Trento Centre for Local Development) for making the paper publication ready.

Executive summary

Data on online job postings represent an important source of information for local labour markets.

Commercial data providers, such as Lightcast, collect real-time job postings and extract metadata, such as occupations, sector, skill needs and the location of a job. Therefore, such data can provide timely and detailed information on labour and skills demand in local economies. However, analysts need to be aware about the limits of online postings for labour market analysis, because there will be differences across places, sectors, occupations and over time in the use of online tools in the job market.

This paper offers an international assessment of the representativeness of web-scraped vacancy data from one provider (Lightcast) for European countries. Detailed inquiries are made with national data for Austria, Belgium, Germany, the Netherlands, Portugal, Spain and Sweden, and with Eurostat hosted data for Bulgaria, Hungary, and Romania. The paper considers distributions across occupational categories, industrial sectors and regions between 2019 and 2022. It does so in comparison with official vacancy data from national surveys on vacancies and job openings and from records of public employment services. A companion paper covers Lightcast data for Australia, Canada, the United Kingdom and the United States for 2015-22. For these countries, a longer time period is available, because Lightcast and its predecessors started in English speaking countries, resulting in datasets that are more developed relative to those of European countries.

While the number of online job postings has increase over time in European countries, large differences with statistical or Public Employment Services (PES) sources between countries, regions, sectors and occupations remain. Lightcast data gathering is under constant development, increasing the number of vacancies that are gathered. The ratio of Lightcast vacancies over vacancy numbers from national sources has increased across most European countries. Large cross-country differences in the ratio of Lightcast online job ads with other sources are also apparent. Using national sources that have suitable regional or sectoral data also highlights the differences in coverage within countries. For instance, urban regions and cities may have a larger share of jobs advertised online compared with remote areas. Lightcast data tend to overrepresent cities in some countries, but in others the data underrepresent cities.

The statistics presented in this paper do not directly test whether one data source is more representative of the labour market than the other. While online job postings can have some limitations in covering the entire labour market, so do other sources. For instance, data from PES may be not be representative of the entire labour market if they only include vacancies that firms submit voluntarily to the agencies. Survey based measures also suffer from limitations since the implementation of surveys differs across countries. Therefore, the statistics are indicative only of a similarity between sources.

Regional coverage of Lightcast data appears to match best in countries where PES have relatively well-developed online tools for jobseekers and employers. The match between Lightcast data and administrative sources can be weak in some countries. At the most aggregate level, Lightcast captures the differences between countries, sectors and occupations relatively well. PES data of the Netherlands, Germany and Sweden provide a greater similarity across regions and sectors with Lightcast data relative to the survey-based data from the same countries. Hence, where PES agencies have systems in place to

collect and advertise vacancies broadly, their public data also matches well with the data that Lightcast collects. Countries that lack any detailed vacancy data from official sources prevents establishing whether data on online job postings is representative of the labour market.

These findings have implications for labour market analysts interested in using data on online job postings from private sector companies for regional, sectoral or occupational analysis of labour markets in European countries. For instance, it cannot be assumed that the data coverage is the same across countries, while strong growth of collected ads over time may differ between countries, sectors and occupations. To assess whether such issues are present, analysts can consider to

- Benchmark their sample against other vacancy sources at an aggregation relevant for their case,
- Consider alternative sources, such as PES agencies, if available and it suits the research purpose,
- Apply statistical techniques, e.g. data weighting, to correct for under- and overrepresentation of some dimensions of online job postings data.

1 Introduction

Data on online job postings represents an important source of information for local labour markets.

Online job postings can be collected from numerous websites, ranging from public employment services (PES) to commercial recruitment firms to individual company websites. Often, job postings specify precisely where a job is located. In that case, the data can provide timely and detailed information on labour and skills demand in local economies.

Regional trends in hiring can contain important information for policymakers about economic performance as well as the jobs and skills in demand.

Growth in the number of vacancies in a labour market can indicate firms' positive outlook for the future and growth of business activity. Detailed information on vacancies can also give insights about the trends in the demand across sectors and occupations. Individual job ads will also include information on specific education and skills requirements. For instance, indicators for skills demand in local labour markets can be used to track local policy implications of technological, demographic and environmental transitions (OECD, 2020^[1]; OECD, 2020^[2]; OECD, 2019^[3]). To draw accurate conclusions on such demand trends, knowledge of the representativeness of online job-postings data is essential.

Many countries lack statistics on labour demand that are sufficiently up-to-date and disaggregated across regions, sectors and occupations.

The principal source of job openings and vacancies in many countries are surveys among employers run by national statistical organisations (NSO). Surveys are typically run a few times per year, for instance once every quarter. Surveys can be expensive to run and time intensive, while data from surveys require processing, leading to further delay between the reference period of the survey and the date of publication. The statistics are aggregated to preserve confidentiality of survey respondents while ascertaining the representativeness across subgroups in the population. Consequently, vacancy numbers derived from surveys, while potentially representative at the national level and internationally comparable, can be delayed and lack important subnational detail. Some Public Employment Services (PES) make data on their vacancies database available, allowing for a higher frequency through automatic processes. However, data from PES offices may be limited to voluntary vacancy submissions and hence not cover the entire labour market.

Workers increasingly use online tools in their job search

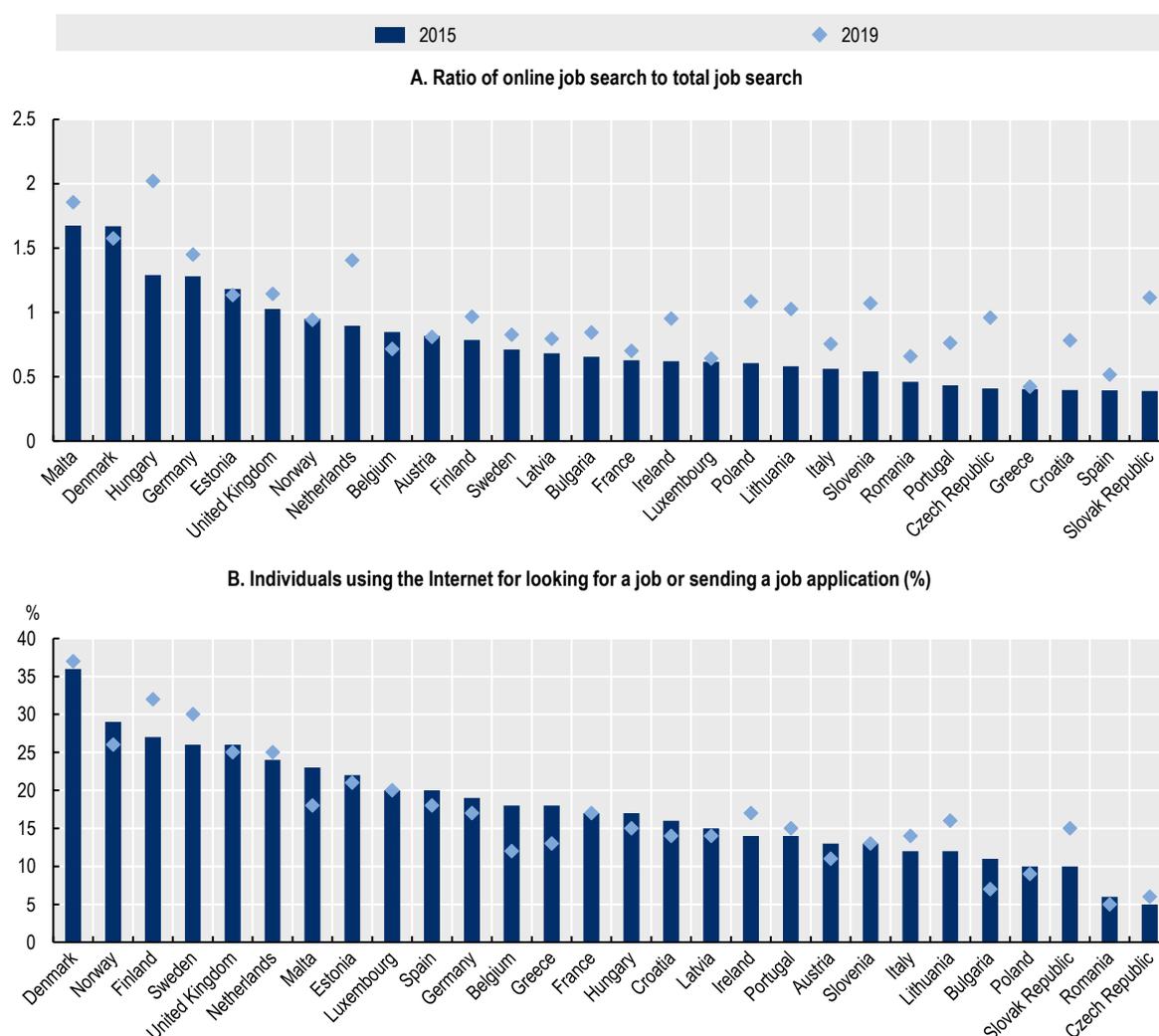
Online tools for advertising and finding jobs have grown more important for employers and job seekers.

Online job searching has lowered the information costs and market frictions, which in turn can shorten job-search periods and improve the matching between employers and employees (Autor, 2001^[4]; Bhuller, Kostøl and Vigtel, 2021^[5]; Gürtzgen et al., 2021^[6]). However, it may also keep job matching rates unchanged as employers can post more vacancies and each online vacancy draws more applicants (Gürtzgen et al., 2021^[7]). Already in 2008-2009, unemployed workers were three times more likely to find a job in the United States using online tools rather than more traditional methods such as newspapers or referrals (Kuhn and Mansour, 2014^[8]).

In Europe, the use of the Internet to look for and apply for a job has increased in 24 out of 29 countries between 2015 and 2019. Figure 1 Panel A shows the share of survey respondents who use the Internet for job search and application over the share of all people who indicated having looked for a

job recently. It is an estimate of the intensity of Internet use for job search and applications across countries over that period. In 24 out of 29 countries, the intensity of Internet use for job search and application increased between 2015 and 2019. In 11 countries, the rate of online job search is higher than the estimate of total job search, indicating people may use online job search for more general information on their career options. Figure 1 Panel B presents the share of people who indicated using online tools, i.e., the numerator of Panel A in 2015 and 2019. Substantial differences between countries are visible, with the use of the Internet for job search activities ranging from over 35% of respondents aged 16-74 in Denmark to 5% in Romania. While the percentage has increased in about half the countries over the five-year period, these rates may also be affected by the economic cycle, which is accounted for in Panel A.

Figure 1. Using internet for job applications is increasing across most EU countries



Notes: Panel A: “online job search” is the percentage of all workers using the Internet for looking for a job or sending a job application) in the last three months. Total job search is the share of workers who i) look for another job while employed, or ii) seek work while not in employment, in the last four weeks, multiplied by three to match with period of online job search. Panel B: The numerator of panel A. The multiplication of total job search by three assumes that the group of people looking for jobs between month two and three of a quarter is separate from those looking for jobs in month one of the same quarter. In reality, an overlap in this population is likely, and so the multiplication should be lower than three. The calculated ratio in Panel A therefore represents a lower-bound estimate.

Source: OECD calculations based on Eurostat table TIN00102 and European Labour Force Survey.

Online job searching provides a new data source for labour market analysis

Web-scraped data from online job postings can provide further insights on the trends in labour demand and the skills needed across regions, sectors and occupations. Various private sector data firms have developed datasets based on online job ads. Such data is based on web-scraping algorithms across many online job portals and company websites. In countries or regions where virtually all jobs are posted online, such data can potentially cover the entire labour market. Automated processes make sure that data can be made available with little delay, while the full text of a job ad can be processed to extract information on various relevant indicators ranging from sector and occupational information, to salary ranges, and skills and work experience requirements.

Lightcast is a private sector data provider whose data has been used in labour market research across a range of topics. Lightcast (formerly known as EMSI Burning Glass and Burning Glass Technologies) data have been used to understand the adoption of new skills (Hershbein and Kahn, 2016^[9]) and technologies (Goldfarb, Taska and Teodoridis, 2020^[10]) in the labour market as well as assessing changes during recessions (Modestino AS, 2016^[11]), such as the COVID-19 pandemic (Tsvetkova, Grabner and Vermeulen, 2020^[12]; Grabner and Tsvetkova, 2022^[13]), among others (Cammeraat and Squicciarini, 2021^[14]). Lightcast is not the only provider of web-scraped job postings. The companies Indeed, Textkernel, and Jobdigger also collect such data for various countries. Moreover, Eurostat is establishing its own implementation and data collection (Pouliakas, 2021^[15]).

Analysts need to be aware of the limits of using online postings in analysis across regions, sectors and occupations for each country and over time. Data providers rely on the development of online platforms for job searching and job postings. Therefore, data of online job postings may not always be fully representative of the labour market across places, sectors, occupations and over time.

This paper assesses the comparability and validity between Lightcast and other data sources for European countries. The other data sources are Eurostat statistics on vacancies for all European countries, and national statistical sources and national public employment services from seven European countries. The first contribution of this paper is to provide evidence on the validity of Lightcast data for regional job market analysis.¹ The analysis also compares the number of vacancies reported by the available administrative data against the Lightcast data for occupations (1 digit) and sectors (1 digit) across countries and time. For some countries, this paper can incorporate more disaggregated sectors and occupations. Secondly, it catalogues which national sources are available for direct comparison, thereby indicating where online job postings might complement administrative sources.

Data on individual online job postings from Lightcast for European countries complement data for English-speaking countries. For European countries, data are available since 2018, while for the US, data are available since 2007 and for Canada, the UK, Australia and Singapore, data are available since 2012. This paper focuses on the European countries, while a companion paper focuses on the English-speaking countries (Tsvetkova et al., forthcoming^[16]).

The remainder of the paper is structured as follows. The next section provides background on the data sources and methodology for benchmarking the data. This is followed by results that depict the representativeness of Lightcast data across countries, regions within countries, sectors and occupations. The paper concludes with a discussion on the findings and a few suggestions for labour market analysts interested in using data on online job-postings.

¹ Few papers have benchmarked the European data of Lightcast. Lovaglio, Mezzanzanica and Colombo (2019^[21]) present an early version of web-scraped Italian job postings from 2013 to 2018, which are compared with survey data from the Italian national statistics office.

2 Data and methodology

This section briefly describes the three types of data sources used in this paper. The three data types are

1. Lightcast online job postings by country,
2. Eurostat aggregated job vacancy data by country, and
3. National sources from selected European countries, which include
 - a. National Statistical Offices (NSO) and
 - b. National public employment services (PES).

This section spells out the main characteristics of the source and highlights how differences in these methodologies could lead to differences in the number of counted vacancies and the distribution across regions, sectors and occupations. Moreover, the representativeness of the source for the entire labour market is discussed.

Lightcast online job postings data

This paper uses data on online job postings as compiled by Lightcast, a commercial labour market and employment analytics firm. The company collects postings through a continuous web scraping algorithm over many online sources for job ads. Lightcast data provide individual job ads at a daily frequency, while its algorithms aim to extract information on the location by city (local administrative units) and/or NUTS3 region, industry, and occupation.² For each job ad, the company's algorithms also extract a list of the required skills and the type of contract. The company aims to remove duplicate ads that originate from the same ad posted on multiple platforms and it tracks ads that remain online for some time. Whether a firm uses one job ad to hire multiple persons is unknown. Lightcast depends on the timeliness of its sources, but job ads are often not removed immediately after a vacancy is filled. Therefore the company implements further data cleaning to limit the duration that it records a vacancy as unfilled. An in-depth discussion on the advantages and disadvantages of Lightcast data is provided in (Tsvetkova et al., forthcoming_[16]). The data covers all 27 European Union member states. In this paper, the data are aggregated across time, sector, occupation and region to match the lowest common denominator for the administrative sources.

Eurostat data

Eurostat gathers and publishes harmonised statistics from the EU Member States in partnership with the NSOs. Eurostat defines a job vacancy as a “paid post that is newly created, unoccupied or about to be vacant, for which the employer is taking active steps [to find suitable a candidate], and [which the employer intends to fill]”.³ The data are categorised by economic activity in accordance with the Statistical

² NUTS3 follows the EU nomenclature for small regions, and is largely equivalent to the OECD definition of TL3 regions for European countries.

³ See https://ec.europa.eu/eurostat/cache/metadata/en/jvs_esms.htm

Classification of Economic Activities in the European Community (NACE Rev. 2), and by occupation, using the International Standard Classification of Occupations (ISCO). However, the data are only available at a country level and countries such as France and Italy are not fully covered. The data are currently available with a delay of approximately one quarter. Napierala, Kvetan and Branka (2022^[17]) further note that the collection methodology differs across some countries, which reduces the homogeneity of the data for comparison across countries.

National sources

National statistical and administrative sources from seven European countries are used in this paper to complement the Eurostat data. Relative to the Eurostat statistics, national sources may provide greater detail for the respective country, for instance through additional regional information, or more detailed sectoral or occupational breakdowns. This paper considers seven European countries: Austria, Belgium, Germany, the Netherlands, Portugal, Spain and Sweden. The sources of the data are provided in Annex A.

National official sources are typically based on surveys run by NSOs or through vacancies that are gathered by PES. Table 1 provides an overview of the sources used in this paper. The survey-based data may have a long time series, but tend to be infrequent, e.g., quarterly, and publicly available tables provide a limited disaggregation over subnational, sectoral, and occupational dimensions. These limitations are often driven by the limited sample size of the underlying surveys. Published statistics based on survey-based data still need to be statistically representative, and comply with confidentiality requirements.

Table 1. Overview of vacancy data from national statistical sources used in this paper

Country	Regional dimension	Sectoral dimension	Occupational dimension	Time	Organisation	Source
Austria	NUTS2	2 digit	Non-ISCO 2 digit	Monthly	PES	AMS
Belgium	NUTS1	1 digit	-	Yearly	NSO	Statistics Belgium
Germany	Labour agency areas	1 digit	-	Monthly	PES	Bundesagentur für Arbeit (BfA)
The Netherlands	Municipal	Custom sectors	ISCO 4 digits	Weekly	PES	Open data match of UWV
Portugal	NUTS1	1 digit	-	Since 2010, quarterly	NSO	Statistics Portugal
Spain	NUTS1	1 digit	-	Since 2013, quarterly	NSO	Statistics Spain
Sweden	NUTS3	macro sector	-		NSO	Statistics Sweden
	NUTS3	-	-	Since 2020, monthly	PES	Arbetsförmedlingen

Notes: PES: Public Employment Agencies. NSO: National Statistical Offices.

Source: Authors' elaborations

The increased digitalisation of PES has also led to the development of additional public data sources. PES in several countries provide internet-based job-matching tools for job seekers and employers. This is part of a wider development of PES providing digital services to job seekers, which was further boosted during the COVID-19 pandemic (OECD, 2021^[18]). PES databases of vacancies provide a readily available source on the state of labour demand using largely non-confidential information. However, PES sources are not always representative of the demand for workers in the entire labour market. Private firms are often not obliged to post their vacancies with PES. It is likely that PES collect vacancies more tailored towards certain occupational groups and sectors, for instance technical professions, clerical

support and trades workers, thereby underrepresenting managerial professions. In some countries, PES data are augmented with vacancies available through private recruitment agencies, which can potentially mitigate part of such biases and provide a better representation of the labour market. Nevertheless, *a priori* it cannot be assumed that vacancy data from all national PES are representative of their labour markets.

Job postings, open vacancies and surveys

This paper compares three types of sources of vacancies, but each source counts vacancies in different ways. The different ways of counting vacancies is likely to lead to conceptual and statistical differences among sources. Lightcast counts new job postings, PES services provide the number of open vacancies and business surveys report the number of vacancies employers have recently opened. Table 2 summarises the three approaches.

Lightcast provides new online job postings, which can be aggregated across time, regions, sectors and occupations to create an indicator of the number of new vacancies. Since vacancies can remain open for some time, vacancies that are already registered are ignored. When a vacancy remains open for several months, Lightcast assumes that further hiring is taking place and recounts the vacancy as a new job posting.

PES services report open vacancies, and the data used in this paper uses those aggregated across sector, occupation or region. This can be done by taking all open vacancies at a specific date, or aggregating all vacancies that were open in the reporting period. The difference with the number of new job postings is that vacancies that remain open for an extended period can be recounted from period to period. For instance, if a vacancy remains open for one month, and the PES services report a weekly number of open vacancies, the same vacancy may be counted four times, but Lightcast would count the same vacancy only once.

Table 2. Three sources, three ways to count vacancies

Source	Vacancies indicator	Description	Reference
Lightcast	New online job postings	Newly recorded job vacancy on website. Vacancies that are reported at multiple websites are removed using best efforts. Vacancies that are still open for more than three months are recounted as a new opening.	
PES	Current open vacancies or open vacancy over a period	Current vacancies in PES database, at time of reporting or aggregating over a period. Vacancies that are still open from period to period are recounted as open vacancies.	The Dutch source includes both registered vacancies at the PES (UWV) and scraped data from other job platforms. ^a Data from the German PES (BfA) includes only registered vacancies. ^b
NSO, following Eurostat	Current open vacancies	A newly opened, to be opened, or unoccupied paid post, for which the employer is taking active measures to recruit, for instance through a job posting, and intends to fill. European NSOs will follow the same definition but may use a variety of sources and sampling methods.	https://ec.europa.eu/eurostat/cache/meta/data/en/jvs_esms.htm

Notes: a. UWV, Specificatie UWV open match data v1.3, 3 march 2022, Available at :

<https://data.overheid.nl/sites/default/files/dataset/84e76ca2-f91b-427d-918b-6a6d8299b3d3/resources/20220329%20Specificatie%20%E2%80%98open%20match%20data%E2%80%99%20UWV%20WERKbedrijf%20V13.pdf>; b. Bundesagentur für Arbeit, Statistik, Statistik über gemeldete Arbeitsstellen, July 2020.

<https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Definitionen/Generische-Publikationen/Kurzinformation-Gemeldete-Arbeitsstellen.pdf>

Source: OECD elaboration.

National statistical agencies use a survey to ask firms about their current hiring and open vacancies. As surveys are conducted infrequently, e.g., once per quarter, and the question may only relate to the current situation, not total of firm hiring over the quarter, the vacancy count is likely to be lower than the other two sources that provide a much higher frequency.

The three ways of collecting vacancy data will lead to different levels of vacancies, and may affect the dimensions of the data (time, region, sector, occupation) in different ways. Differences in the count of vacancies between sources may occur if sectors or occupations have different hiring patterns. For example, some sectors are more likely to hire within a week of the initial posting, while other sectors frequently require several months. Such differences in hiring patterns would lead to differences in the relative size of vacancies between sources. With regional specialisation in specific occupations or sectors, such differences in hiring patterns will also affect aggregate patterns across regions.

Methodology

A simple ratio of the number of vacancies across sources can provide a first indicator of their comparability. A perfect match in the number of job postings across sectors or occupations and across time and regions would lead to ratios that are close to one. This would be a hypothetical situation where all vacancies are posted online and both sources draw on the same information and count vacancies in the same way. However, not all vacancies are posted online, and jobs for some sectors or occupations are more likely to be posted online. Moreover, a difference in the collection method and its frequency could create substantial differences in the number of vacancies between two sources. For instance, a national source may be based on a quarterly survey question asking employers their vacancy over a short period.

The ratios are calculated on yearly sums of vacancies, even if most countries provide data at a higher frequency. The yearly sums help to present the results over the four-year period of 2019-22, and to understand the development of comparability of online job postings data with administrative sources across years. However, the yearly sums of vacancies may have a lower comparability relative to quarterly sums, as suggested by Tsvetkova et al. (forthcoming_[16]) based on analysis for Australia, Canada, the United Kingdom and the United States.

In addition to the ratios, regression-based conditional correlations can be used when data sources allow. When the discrepancies and biases are structural to one of the dimensions in the data, e.g., time, occupation, or sectors, conditional correlations may partly correct for these. Conditional correlations are estimated using simple OLS regressions on disaggregated data with various combinations of fixed effects. These statistics are calculated on data sources that provide the greatest disaggregation, e.g., with sector by region information or with a high detail in the occupational codes.

The statistics presented in this paper do not directly test whether one data source is more representative of the labour market than the other. Data from Eurostat and the surveys of national statistical organisations may be most representative of the job market overall. The survey methods aim to provide an accurate estimate of job demand, covering all sectors, occupations. Nevertheless, different implementation methodologies across countries can indicate that such methods may not always fully reflect the underlying job market either (Napierala, Kvetan and Branka, 2022_[17]). PES sources may not be representative since their main purpose is to help jobseekers. Therefore, this paper treats the calculated statistics as a measure of comparability of the different data sources rather than making a statement of representativeness of data source of the underlying population.

3 Results

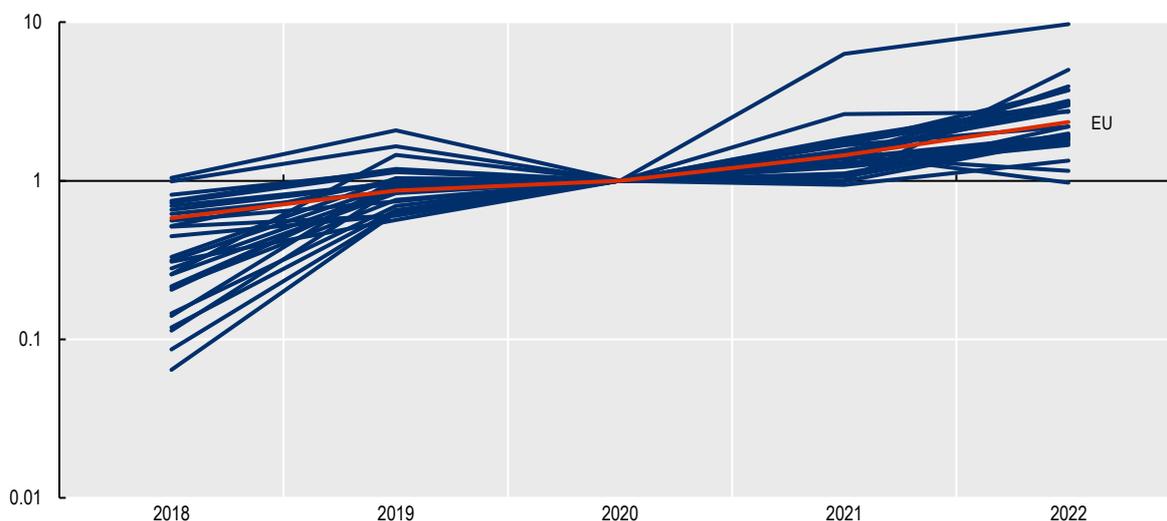
The results are discussed in three parts. First, a European overview of Lightcast is provided and a first comparison against similar counts from Eurostat is used to demonstrate how Lightcast data compares with vacancy data from Eurostat. Second, Lightcast data is assessed for its regional coverage using a variety of national sources and from Eurostat for several countries. The comparison is repeated for a sectoral and occupational overview. Finally, conditional correlations are presented for some countries where data is sufficiently disaggregated over time, region and either sector or occupation.

The coverage of Lightcast data across countries and time

Lightcast data for European countries starts in 2018, but for many countries a large increase in vacancies between 2018-19 indicates data collection is under continuous development. The data between 2018 and 2019 underwent strong development (Figure 2). Several countries saw the number of vacancies between 2018 and 2019 grow more than 10-fold. For the whole of the EU, the number of vacancies increased by almost 50%.

Figure 2. Lightcast online job postings rise rapidly between 2018 and 2019 for many countries

Ratio of the number of vacancies from Lightcast over the vacancies from Eurostat, by country, log scale, 2019=1.



Note: Each line represents a separate country, each is normalised to the number vacancies reported for 2020. The highlighted EU line is the average of the other lines weighted by underlying number of vacancies.

Source: OECD calculations based on Lightcast data.

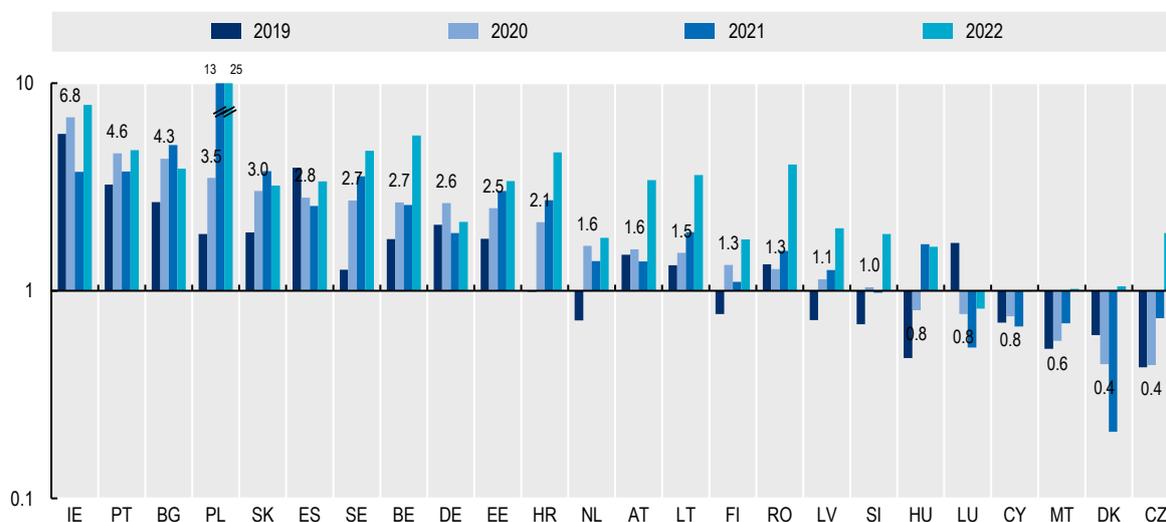
The decline between 2019 and 2020 observed in some countries can be related to the COVID-19 pandemic. Between 2020 and 2021, some countries show flat lines while others experience substantial increases. The EU overall experienced a growth of 15% over 2020-21. However, between 2021-22, the number of Lightcast vacancies across EU countries increased by 63%. This likely reflects increased data gathering on top of the general increased labour demand during economic recovery from the economic disruption of the COVID-19 pandemic.

The total vacancies counted in Lightcast by country and year can be compared with the total number of vacancies reported by Eurostat. The resulting ratios provide a first indication on the comparability between the sources. Box 1 provides a further explainer on how to interpret the graphs presented in this paper.

The ratios for 2020 vary between 6.8 in Ireland to 0.4 in Czechia, indicating that the two sources do not follow the same distribution. The number of Lightcast vacancies relative to those reported by Eurostat in 2020 is more than four times higher in three countries and more than double in eight more countries (Figure 3). In six countries, the number of vacancies in Lightcast is lower than those reported by Eurostat for 2020. The numbers would be expected to be more similar across countries if Lightcast and Eurostat capture national labour markets equally well, even if the two sources count vacancies differently. Moreover, there is little evidence that the distributions are converging with time. The countries with already high bars in 2020 see generally further increases for 2021. The 2021 and 2022 data for Poland appears as an additional outlier.

Figure 3. Lightcast and Eurostat do not have equal distributions across countries

Ratio of Lightcast online job postings over Eurostat vacancies across country and year.



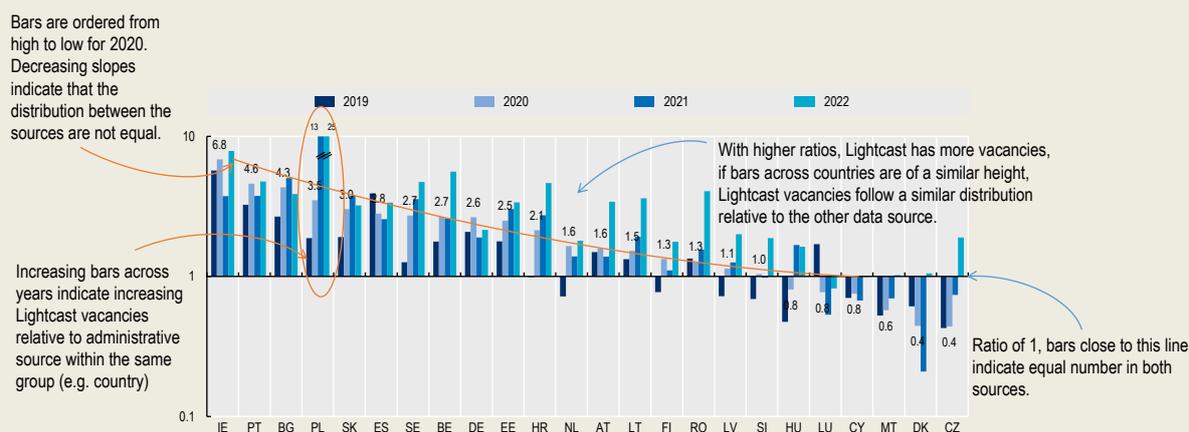
Note: numbers above bars refer to 2020.

Source: OECD calculations based on Lightcast and Eurostat table jvs_q_nace2.

Box 1. A visual method to compare how Lightcast vacancies compare to other sources across groups and time.

A visual representation of the comparability between the two sources can be provided by the ratio of the number of vacancies in Lightcast data relative to the official sources, by subgroups and year. Figure 4 provides some guidance for the interpretation of the graphs in this paper.

Figure 4. Interpretation of the bar graphs



Source: OECD elaboration. See Figure 3.

Three main aspects can be read from the graph. First, the height of the bars indicates how many more (or fewer) vacancies are recorded in Lightcast data relative to the other sources. Bars close to the value one, the horizontal axis, provide a reference value for bars that indicate that both sources have approximately similar numbers. If bars are much higher than the reference value one, then there can still be a high similarity in the distribution between the two sources. This would be the case if all bars have a similar height.

Second, A decreasing pattern of bars indicate different distributions across the groups in the two sources. The bars are sorted from the highest to the lowest for a reference year, in this paper this will be 2020. The order of the bars gives a decreasing pattern across groups. The stronger the pattern the lower the correlation between the two sources. The figure illustrates which groups are driving the discrepancy, for instance by showing big differences in the height of the bars at the left and right end of the order. In Figure 4, the groups are countries, but the data could be also split by different groups, such as sector or occupation.

Third, the coloured bars provide a differentiation by year. For a given group, an increasing length of upward bars across the years, and a decreasing length of downward bars, indicate that the number of vacancies in Lightcast increases within the group. This could be due to changing algorithms of Lightcast or a general increase in job ads available from online sources, but it does not automatically mean that the Lightcast becomes more closely aligned to the official sources. An increasing alignment by time would be represented by convergence of bars to the same level.

The coverage of Lightcast data across subnational regions

Regional disaggregation for some countries is provided by some NSOs and PES. Belgium vacancy data is provided by the Statbel, the Belgian statistical office. For Hungary, Bulgaria and Romania, regional vacancy statistics are provided by Eurostat and for Austria, Germany and the Netherlands such data are provided by the national PES.⁴ For Sweden, data from both the NSO and the PES are included. For each country-source, the ratio of the total vacancies in Lightcast relative to the national source is calculated at the smallest possible regional aggregation.

For most countries, the ratio of Lightcast vacancies to the national source varies strongly across regions (Figure 5). In Austria and Germany (panels A and C), as well as the three European countries covered by Eurostat (panel B), the ratio varies from values below one to far above one. Values below one indicate an undercount of Lightcast relative to the alternative source, and values above one indicate that there are more vacancies in the Lightcast data than the national source. For Sweden (based on NSO) and Belgium (panel A), Lightcast always indicates a far greater number of vacancies.

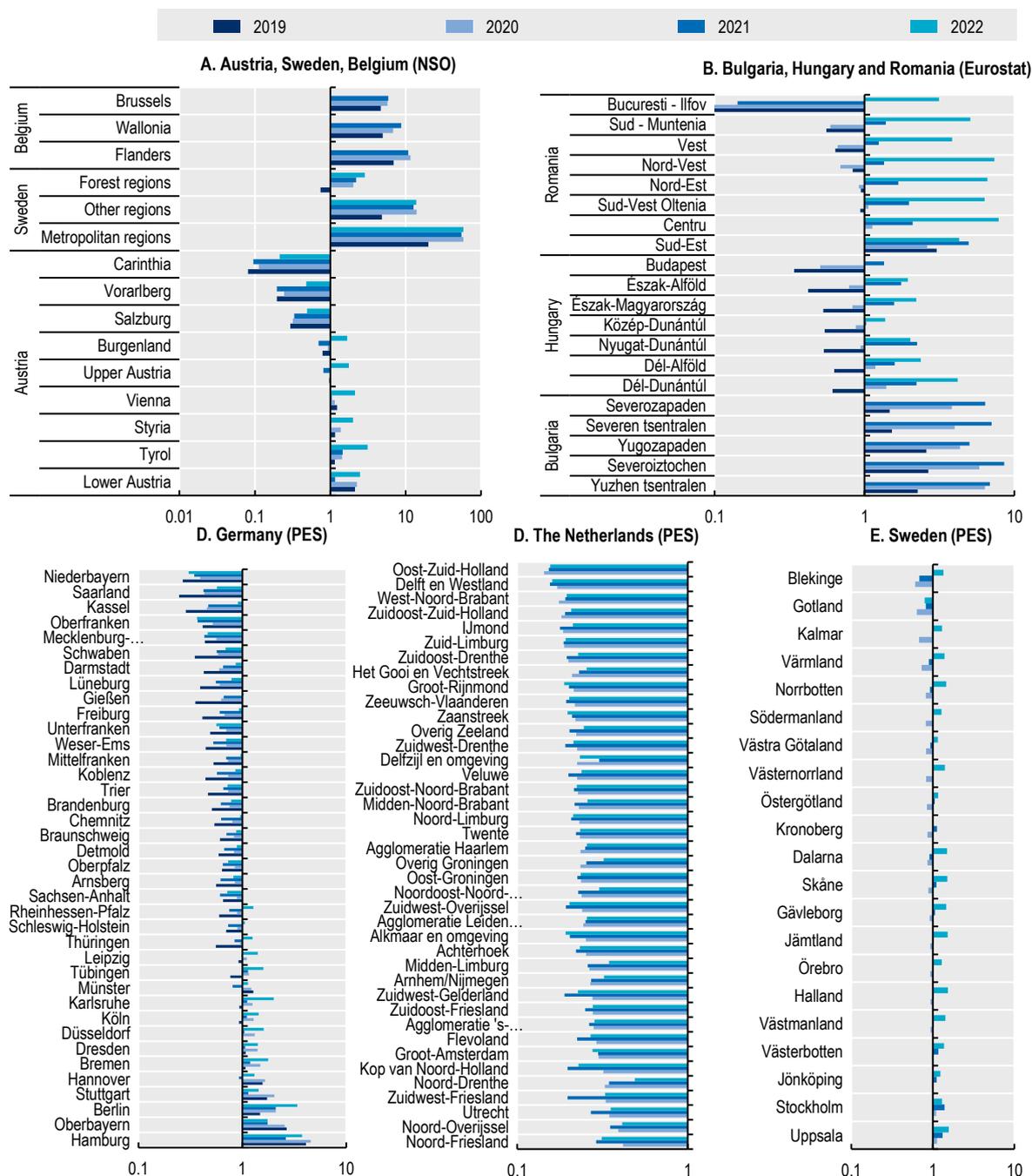
Data of the public employment services can closely mirror the numbers of Lightcast. The bars across the Dutch regions (panel D) are of similar height for each of the three years, indicating that Lightcast and the Dutch source provide a very similar distribution across regions for all three years. Moreover, the bars for 2021 appear to correct for some regional discrepancy in 2020, with regions that have the longest and shorter bars in 2019 converging towards the levels seen in the middle. The ratio is consistently below one, indicating that the number of vacancies quoted by the public employment agency in the Netherlands is well above those that are provided by Lightcast. This may be due to Dutch PES reporting the stock of open vacancies rather than new job postings (See Table 2). Results based on the Swedish PES data (panel E) provide ratios close to one for all regions across all of the three years in comparison to the other country cases.

Online job postings are not always highest in the most urban region of a country. The uptake of the use of Internet tools tends to be higher in cities (OECD, 2020_[19]). Therefore, it may be expected that urban areas are better covered in Lightcast resulting in larger ratios for these areas. This pattern is not present in all countries. In Belgium (panel A), Flanders has the highest, but Brussels has the lowest ratio. Bulgaria has a high ratio for Yugoispaden (containing the capital Sofia), but Budapest and Bucharest have the lowest ratio for Hungary and Romania respectively (panel B). For Germany (panel C), large cities have the highest ratios, such as Hamburg and Berlin, but in the Netherlands (panel D), the NUTS3 areas that contain the main cities (Groot-Amsterdam for Amsterdam, Groot-Rijnmond for Rotterdam, Agglomeratie 's-Gravenhage for The Hague and Utrecht) are not all ranked towards the high-end of the graph. In Sweden, using PES data (panel E), the Lightcast data appears to have more vacancies for the Stockholm and Uppsala, but the difference across the regions is minor relative to the variation seen in the other panels.

⁴ Regional information is also provided by CBS, the national statistical office of the Netherlands, but at NUTS2, whereas data from UWV, the Dutch PES allows for a comparison at NUTS3. Therefore, the text presents the data for the Netherlands at NUTS3.

Figure 5. The number of online job postings relative to national sources varies across regions for most countries

The ratio of Lightcast vacancies over national source, by region and year.



Note: For Belgium, Germany, Bulgaria, Romania and Hungary the regional breakdown is at NUTS2. For the Netherlands this is at NUTS3. Sources in brackets, NSO for national statistics office. In panel A, Sweden is based on survey statistics of Statistics Sweden, which provides a custom aggregation based on NUTS3 regions and was replicated in Lightcast data. In panel E data is based on Swedish public employment agency data, available at NUTS3. In panel D, German data is available at more detailed region of administration of its public employment services, which were mapped to NUTS2 regions, but cannot be mapped consistently to NUTS3 regions.

Source: OECD calculations based on Lightcast and national data: AMS (Austria), BfA (Germany), UWV (The Netherlands), Statistics Belgium, Statistics Sweden, Arbetsförmedlingen (Sweden) and Eurostat

Sectoral coverage

National sectoral coverage can be compared using the NACE2 classification, presented in Table 3, for most of the included countries. For Germany and Portugal, sectors were combined to match the national sources with Lightcast. Austrian, German and Swedish data is based on PES data, and for the Netherlands, Portugal and Spain on national statistics.

Table 3. Sectors according to NACE classification

	Sector		Sector
A	Agriculture, forestry and fishing	K	Financial and insurance activities
B	Mining and quarrying	L	Real estate activities
C	Manufacturing	M	Professional, scientific and technical activities
D	Electricity, gas, steam and air conditioning supply	N	Administrative and support service activities
E	Water supply; sewerage, waste management and remediation activities	O	Public administration and defence; compulsory social security
F	Construction	P	Education
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	Q	Human health and social work activities
H	Transportation and storage	R	Arts, entertainment and recreation
I	Accommodation and food service activities	S	Other service activities
J	Information and communication		

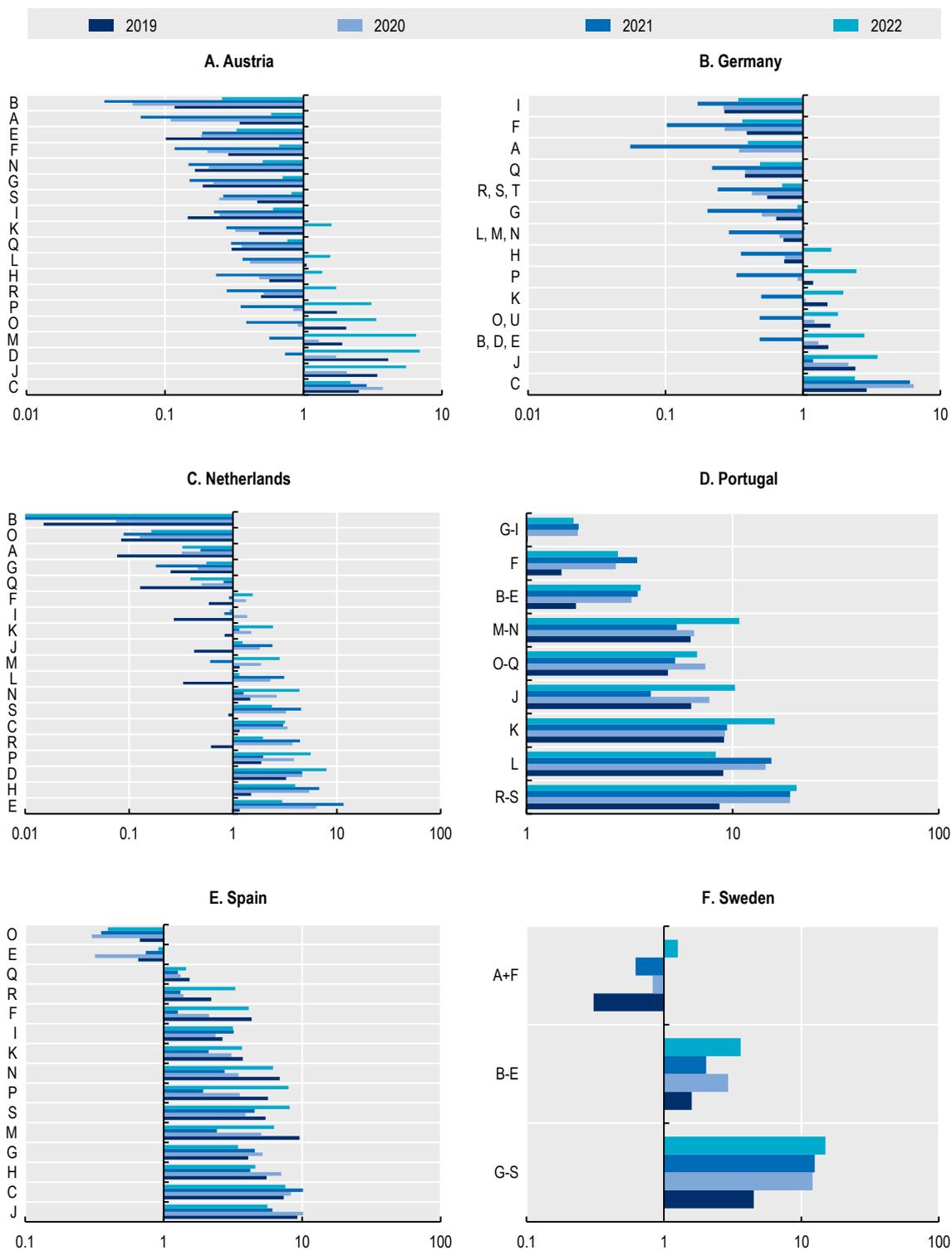
Source: https://www.oecd.org/sti/ind/STANI4_Industries_ENG.pdf

In each of the included countries, Lightcast data follows a different distribution over sectors relative to the corresponding national source. Figure 6 shows that in five of the six countries, the ratio varies from below one, indicative of higher vacancies in the national source, to above one. Manufacturing (C), Utilities (D) and Information and communication (J) have the largest ratio in Austria, Germany, the Netherlands and Spain.

Similar unequal distributions are observed for countries that are covered in both Lightcast and Eurostat datasets. Figure 7 shows that the combined sector that includes manufacturing and utilities (B to E) has the highest ratio in 2020, followed by the professional and administrative services (M and N) and Information and communication (J). However, these rankings change for 2021 and 2022. In 2022, Lightcast presents vastly more vacancies in financial and insurance activities (K), Professional, scientific and technical activities (M), and other services sectors belong to J and R-S.

Figure 6. Sectoral distributions of vacancies between online job postings and national sources

The ratio Lightcast vacancies over national source, by sector and year.

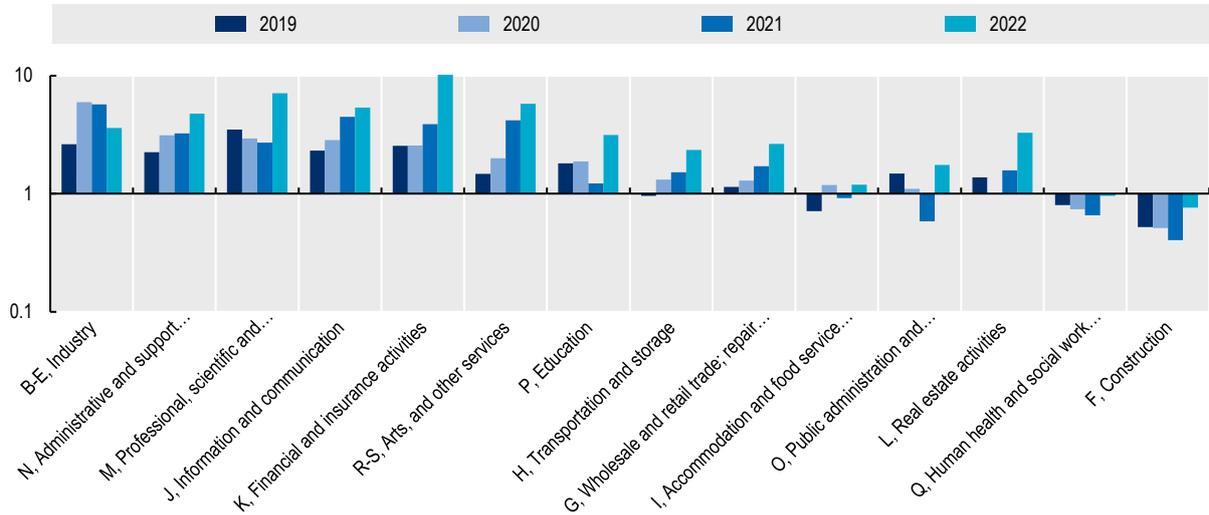


Note: Sector labels omitted for space, see Table 3.

Source: OECD calculations based on Lightcast and Austrian: AMS, Germany: BfA, Netherlands: CBS, Portugal: INE, Spain: INE, Sweden: Statistics Sweden.

Figure 7. The ratio of online job postings over aggregate EU vacancies across sectors varies over time and sector

Share of vacancies across sectors by source, 2019-22.



Note: For full sector labels, see Table 3.

Source: OECD calculations based on Lightcast and Eurostat table jvs_q_nace2.

Occupational coverage

The last dimension of comparison is by occupational groupings. At the most aggregate level, occupations in business services may be more frequently posted online relative to jobs trades and manufacturing occupations. Such biases in online job advertisement across occupational groups could lead to biases in the representativeness of online job postings data.

The occupational coverage follows the ISCO classification. Table 4 provides the 1-digit occupations for references. For Bulgaria, Romania and Hungary, a breakdown at 1-digit is available from Eurostat. Some national sources allow analysis at a greater occupational detail. For the Netherlands, the PES provides 4-digit codes, which are reduced to one digit for the graphical analysis, but the more detailed codes will be further explored below. Austria provides occupational codes that require further aggregation to match the national source to the ISCO codes that are available in Lightcast (see further details in 0). The public data of German PES does not follow ISCO codes for its job market data across occupations.

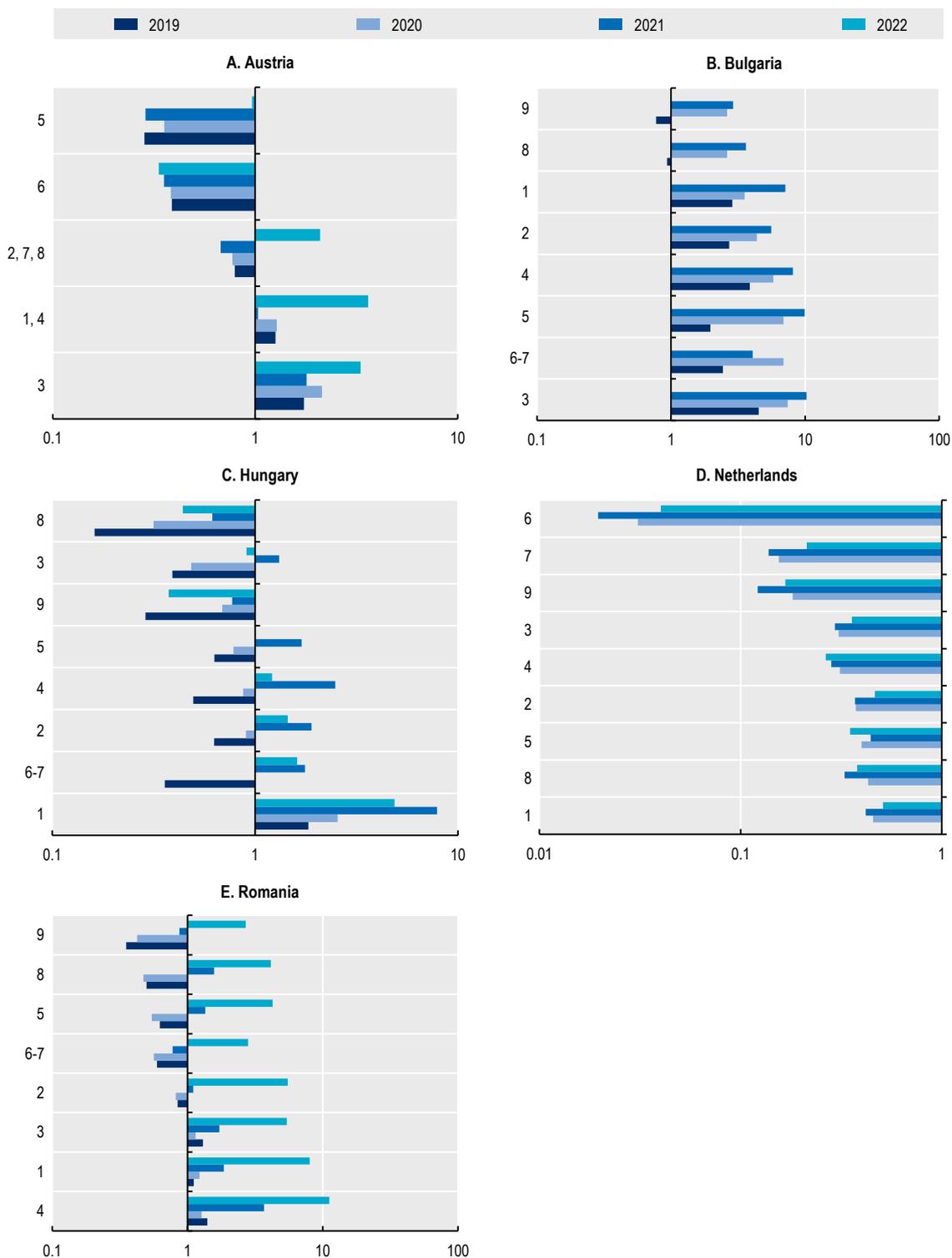
Table 4. ISCO occupations

Code	Title	Code	Title	Code	Title
1	Managers	4	Clerical support workers	7	Craft and related trades workers
2	Professionals	5	Service and sales workers	8	Plant and machine operators and assemblers
3	Technicians and associate professionals	6	skilled agricultural, forestry and fishery workers	9	Elementary occupations

Source: <https://esco.ec.europa.eu/en/classification>

Figure 8. Occupational distributions of vacancies between Lightcast and national sources

Ratios Lightcast vacancies over national sources, by ISCO 1 digit code and year.



Note: 1: Managers; 2: Professionals; 3: Technicians and associate professionals; 4: Clerical support workers; 5: Service and sales workers; 6: skilled agricultural, forestry and fishery workers, 7: Craft and related trades workers, 8: Plant and machine operators and assemblers; 9: Elementary occupations. Occupational groups are combined where necessary to match both sources.
Source: OECD calculations based on Lightcast and Austrian: AMS, Netherlands: UWV, and Eurostat.

In most countries, professional and administrative occupations tend to have a higher ratio of vacancies than others. This result can be product of a selection bias in online job postings for office-based jobs relative to others. In the Netherlands, PES offers far more vacancies for skilled agricultural, forestry and fishery workers than Lightcast, relative to, for instance, management and professional occupations.⁵ The differences in the ratio by the occupational codes contrast with the relative flat distribution provided across regions in Figure 5. The difference suggests that even if some occupational groups are not as well covered in Lightcast relative to PES, this does not directly affect the relative coverage across regions.

Conditional correlations for multi-dimensional vacancy counts

Conditional correlation analysis allows the comparison of vacancy numbers across multiple dimensions at the same time. The results in the previous section indicate that the ratio of the number of online job postings relative to the number of vacancies from official sources can vary greatly across regions, sectors and occupations for a given country. This subsection estimates correlations in the vacancies between two sources after absorbing structural differences across time, regions, and sectors or occupations. This analysis is informative because it can provide insights on the origin of the difference between the sources.

For Austria, Belgium, Germany and the Netherlands, Lightcast data are matched to the national sources over three dimensions simultaneously. For each region and period, vacancies in national sources are available by sector or occupation. A linear regression provides a convenient method to assess whether the number of vacancies in Lightcast matches well with the number of vacancies at that more detailed level of analysis. Box 2 provides further details on the methodology.

Box 2. Regression analysis of vacancies numbers

For each source the following regression model is estimated:

$$\text{LC vacancies}_{r,g,t} = \beta \text{national source vacancies}_{r,g,t} + \beta_0 + a_r + b_g + c_t + \varepsilon_{r,g,t} \quad \text{Equation 1}$$

The national source defines the dimension for the region, r , sector or occupation, g , and time, t . The number of vacancies in Lightcast at the same level of aggregation is then regressed against the number of vacancies in the national source. The regression also allows the addition of control variables for each dimension of the data (also referred to as fixed effects). These are summarised by a_r for region, b_g , for the sector or occupation and c_t for time. The coefficient of interest, β , provides the average number of vacancies from the national sources that is equivalent to a Lightcast vacancy. If the two sources measure the stock of vacancies in the economy accurately, the coefficient β can be expected to be close to one. A coefficient above one indicates that Lightcast has more than one vacancy for each in the national source, and a value below one indicates that the national source has more vacancies for each in Lightcast. The standard errors, robust to heteroscedasticity, provide an indication on whether this estimate is statistically consistent across the observations. The inclusion of the control variables may affect the size of the coefficient and standard error if there is a specific discrepancy across one of the dimensions that the controlling fixed effects absorb.

⁵ The lack of Lightcast job postings for agricultural workers relative to the number of jobs is also noted by (Cammeraat and Squicciarini (2021^[14])).

Across time, region and sector, each vacancy of national sources ranges from 0.12 Lightcast vacancies in Austria, using PES data, to 9.8 Lightcast vacancies in Belgium, using NSO data (Table 5). The results underline that the national sources, and potentially also Lightcast, measure vacancies very differently across countries. Even when only comparing the three countries that are based on PES data, the estimates vary. The coefficients for Austria range between 0.12 to 0.16 (6.3 to 8.3 PES recorded vacancies corresponds one Lightcast vacancy) while the coefficients for Germany vary between 0.70 and 0.84 (1.2 to 1.4 PES vacancies to one Lightcast vacancy) and in the Netherlands between 0.50 to 0.78 (1.3 to 2 PES vacancies to one Lightcast vacancy). The precision of the estimates, indicated by the statistical significance of the coefficient, shows that there is a strong correspondence within each country over the three dimensions of time, region and sector. For each country, the inclusion of the fixed effects for the three dimensions does not affect the coefficients substantially as these remain in the same order of magnitude.

Table 5. A national source vacancy equals between 0.12 to 9.8 vacancies in Lightcast

Linear regression of vacancy numbers in Lightcast over region, time and sector against national sources.

national source	Dependent variable: Lightcast vacancy count					
	(1)	(2)	(3)	(4)	(5)	(6)
a. Austria, public employment services						
Vacancies	0.139***	0.132***	0.136***	0.160***	0.127***	0.121***
	(14.93)	(14.21)	(15.09)	(11.94)	(14.30)	(9.54)
N	25 074	25 074	25 074	25 073	25 074	25 073
R-squared	0.0906	0.115	0.154	0.247	0.181	0.353
b. Belgium, Statistics Belgium						
Vacancies	9.331***	8.659***	9.295***	9.839***	8.598***	8.859***
	(5.04)	(4.46)	(5.12)	(5.83)	(4.54)	(4.60)
N	161	161	161	161	161	161
R-squared	0.463	0.469	0.472	0.667	0.477	0.678
c. Germany, public employment services						
Vacancies	0.761***	0.719***	0.762***	0.839***	0.719***	0.695***
	(35.97)	(36.71)	(36.84)	(22.94)	(37.87)	(18.35)
N	25 004	25 004	25 004	25 004	25 004	25 004
R-squared	0.242	0.303	0.259	0.470	0.319	0.547
d. Netherlands, national statistics office						
Vacancies	0.672***	0.570***	0.623***	0.793***	0.498***	0.532***
	(15.60)	(12.85)	(16.30)	(16.44)	(12.94)	(10.85)
N	4 104	4 104	4 104	4 104	4 104	4 104
R-squared	0.212	0.235	0.338	0.398	0.366	0.541
FE region		yes			yes	yes
FE Time			yes		yes	yes
FE sector				yes		yes

Note: Conditional correlations of the number of vacancies in Lightcast suitably aggregated over region, time and sector definitions to match with the national source. *** p-value<0.01, ** p-value<0.05, * p-value<0.1, based on heteroskedastic robust standard errors. See Box 2 for further estimations details.

Source: OECD calculations based on EBG and national sources, AMS (Austria), Statistics Belgium, AfB (Germany) and CBS (Netherlands)

Lightcast and national sources may provide a relatively better correspondence to each other at higher levels of aggregation than at lower levels. The R-squared indicates for each estimation what share of the variation of Lightcast vacancies is explained by the national vacancies and the fixed effects. For Belgium, this is 46% without the inclusion of fixed effects, for Germany and the Netherlands around 20% and in Austria lower than 10% (see column 1). Belgium has the lowest number of observations due to the small number of regions (three) across sectors and time.

Aggregation over the dimensions may help increase the correlation between Lightcast and national sources. A lower time frequency or higher aggregation over sectors or occupations may average out some discrepancies between the data. The PES source of the Netherlands provides vacancy data by small regions (NUTS3), by week, at the 4-digits occupational codes. Keeping the regional aggregation time-frequency constant, and decreasing the detail of the occupational codes can provide insight into how relevant this approach can be for analysis. Table 6 presents the results.

Table 6. At higher levels of aggregation of occupations, Lightcast and national sources correlate more

Linear regression of vacancy numbers in Lightcast over region, time and occupation against Netherlands PES source.

	Dependent variable: Lightcast vacancy count					
Vacancies from national source	(1)	(2)	(3)	(4)	(5)	(6)
ISCO level 4	0.101*** (49.04)	0.0873*** (44.69)	0.103*** (49.84)	0.111*** (46.14)	0.0890*** (45.48)	0.0766*** (33.77)
N	200 993	200 993	200 993	200 992	200 993	200 992
R-squared	0.156	0.207	0.198	0.352	0.250	0.477
ISCO level 3	0.168*** (37.33)	0.146*** (31.20)	0.172*** (37.97)	0.180*** (37.53)	0.149*** (31.63)	0.142*** (25.22)
N	114 523	114 523	114 523	114 523	114 523	114 523
R-squared	0.270	0.314	0.321	0.424	0.364	0.527
ISCO level 2	0.196*** (44.77)	0.152*** (35.04)	0.202*** (46.69)	0.211*** (44.04)	0.158*** (36.54)	0.156*** (29.70)
N	48 774	48 774	48 774	48 774	48 774	48 774
R-squared	0.335	0.405	0.413	0.460	0.480	0.602
ISCO level 1	0.265*** (33.87)	0.207*** (23.64)	0.281*** (36.54)	0.266*** (34.50)	0.230*** (25.81)	0.212*** (20.98)
N	12 352	12 352	12 352	12 352	12 352	12 352
R-squared	0.452	0.502	0.571	0.511	0.612	0.675
FE region		yes			yes	yes
FE Time			yes		yes	yes
FE Occupation				yes		yes

Note: *** p-value<0.01, ** p-value<0.05, * p-value<0.1, based on heteroskedastic robust standard errors.

Source: OECD calculations based on data from Lightcast and UWV.

The relation between the Dutch PES and Lightcast vacancy numbers strengthens with each decrease in the occupational detail. At the 4-digit occupational codes, each vacancy in the national source relates to 0.101 vacancies in Lightcast, and the national source explains about 16% of the variation in the Lightcast data. At the 3-digit level the coefficient increases to 0.168, and the R-squared increases to 27%. This pattern is repeated in the ISCO 2-digit and 1-digit levels. At 1-digit, the coefficient is 0.265, more than double the coefficient at the 4-digit level, and the R-squared is 45%, three times as large as at the 4-digit level.

Aggregation over the dimension of a data source can aid the correspondence between Lightcast data and national sources. However, the results do not indicate that this solution creates a very close correspondence to the two data sources (comparing coefficients in column 1 across ISCO levels), as 0.265 is still far removed from the theoretical value of one. The inclusion of fixed effects, which potentially absorb differences between the two sources in the frequency of data collection, coverage of sectors or coverage of regions does not increase the coefficient. For instance, at ISCO level 1, the coefficient decreases from 0.265 without controlling for differences between region, time and occupation, to 0.212 when controlling for all three. This suggests that the two sources still track changes in vacancies quite differently after taking into account structural differences across these three dimensions.

Discussion and conclusion

Vacancy data can be key indicator to track local labour markets, but such data is not available for all places from official national sources. National sources primarily consist of regular surveys to employers, typically managed by national statistical offices, and administrative sources coming from public employment agencies. Statistical offices may be limited in providing full breakdowns across sectors and occupations for small regions due to sample sizes of their surveys. Moreover, surveys are typically not run very frequently, e.g. only once per quarter, with some delay in the time from survey to publication of the numbers. Some Public Employment Services (PES) make data on their vacancies database available, allowing for a higher frequency through automatic processes. However, data from PES offices may be limited to voluntary vacancy submissions. In some countries, PES collaborate with private sector recruiters to extend the database of the PES.

Private sector data firms have emerged to address the issue of the limited availability of up-to-date and disaggregated labour demand statistics. Using big data and machine learning technologies, private sector firms scrape online sources to collect job ads and extract harmonised information from the text in job ads. Their data can contain details for individual job ads, including regional, sectoral and occupational indicators along with other relevant details. This detail makes the data potentially suitable for regional and local job market analysis where public sources are not available. Moreover, as data collection and processing occur continuously, the data is available with minimal delay.

This paper provides some indicators on the comparability of the number of online job ads with national sources across regions, sectors and occupations. Since the development of data based on online job postings across Europe is still relatively new, some questions on their coverage of labour markets remain outstanding. For instance, online job ads may be more commonly found in cities than in rural areas, and more often for business administration occupations than for others. This paper documented for seven European countries which public sources on vacancies are available and how these relate to vacancy numbers from Lightcast, a private sector data provider for online job ads.

Data from PES agencies tend to match relatively closely to the data of Lightcast across regions, while survey based data from national statistical offices indicate large differences in vacancy counts across regions, sectors and occupations. Data from national statistical offices may be infrequent (e.g., quarterly) and limited in their disaggregation (e.g., by region, sector or occupation) but expected to be representative of the labour market. Data from PES agencies of Germany, the Netherlands and Sweden suggest that these sources cover territories relatively well, but the detail of sectors and occupations may be lagging.

Several PES agencies across countries are actively developing information systems that provide timely and detailed data on the number of vacancies as well as registered jobseekers. Some PES agencies, for instance in the Netherlands, collaborate with private sector recruitment and employment firms to combine all relevant data into a single source. It is likely that private sector data providers, such as Lightcast, tap in the same sources that are used by PES agencies to provide vacancy statistics.

The lack of vacancy data from national PES offices may be related to job markets that are still transitioning towards online platforms. Not all national PES offices provide yet detailed information on vacancies. One reason for this is that a practice of online job searching and matching is not developed to the same extent in all countries. In this case a data provider, such as Lightcast, may also be limited in its collection of vacancy data. The limitation can then possibly lead to an overrepresentation in vacancies of some places, sectors or occupations, e.g. high-skilled jobs in business services located urban areas.

These findings have implications for labour market analysts interested in using data on online job postings from private sector companies for regional, sectoral or occupational analysis of labour markets in European countries. For instance, it cannot be assumed that the data coverage is the same across countries, while strong growth of collected ads over time may differ between countries, sectors and occupations. To assess whether such issues are present, analysts can consider to:

- Benchmark their sample against other vacancy sources at an aggregation relevant for their case,
- Consider alternative sources, such PES agencies, if available and it suits the research purpose,
- Apply statistical techniques, e.g. data weighting, to correct for under- and overrepresentation of some dimensions of online job postings data.

This paper does not provide specific procedures to calculate weights or other adjustment mechanisms, because the appropriate data procedure will be case specific. Moreover, for some countries it is not clear that either NSO or PES data provide a fully representative overview of local and sectoral labour markets.

References

- Autor, D. (2001), "Wiring the Labor Market", *Journal of Economic Perspectives*, Vol. 15/1, pp. 25-40, <https://doi.org/10.1257/jep.15.1.25>. [4]
- Bhuller, M., A. Kostøl and T. Vigtel (2021), *The Internet, Search Frictions and Aggregate Unemployment*. [5]
- 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>. [14]
- Goldfarb, A., B. Taska and F. Teodoridis (2020), "Artificial Intelligence in Health Care? Evidence from Online Job Postings", *AEA Papers and Proceedings*, Vol. 110, pp. 400-404. [10]
- Grabner, S. and A. Tsvetkova (2022), "Urban labour market resilience during the Covid-19 pandemic: what is the promise of teleworking?", *Regional Studies*, pp. 1-16, <https://doi.org/10.1080/00343404.2022.2042470>. [13]
- Gürtzgen, N. et al. (2021), "Do digital information technologies help unemployed job seekers find a job? Evidence from the broadband internet expansion in Germany", *European Economic Review*, Vol. 132, p. 103657, <https://doi.org/10.1016/j.euroecorev.2021.103657>. [6]
- Gürtzgen, N. et al. (2021), "Does online search improve the match quality of new hires?", *Labour Economics*, Vol. 70, p. 101981, <https://doi.org/10.1016/j.labeco.2021.101981>. [7]
- Hershbein, B. and L. Kahn (2016), *Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings*, National Bureau of Economic Research, Cambridge, MA, <https://doi.org/10.3386/w22762>. [9]
- Kuhn, P. and H. Mansour (2014), "Is Internet Job Search Still Ineffective?", *The Economic Journal*, Vol. 124/581, pp. 1213-1233, <https://doi.org/10.1111/ecoj.12119>. [8]
- Lovaglio, P., M. Mezzanzanica and E. Colombo (2019), "Comparing time series characteristics of official and web job vacancy data", *Quality & Quantity*, Vol. 54/1, pp. 85-98, <https://doi.org/10.1007/s11135-019-00940-3>. [21]
- Modestino AS, S. (2016), "Changes in Employer Skill Requirements Over the Business Cycle", *Labour Economics*, Vol. 41. [11]
- Napierala, J., V. Kvetan and J. Branka (2022), *Assessing the representativeness of online job advertisements*, Publication Office of the European Union, <https://doi.org/10.2801/807500>. [17]
- OECD (2021), *OECD Employment Outlook 2021: Navigating the COVID-19 Crisis and Recovery*, OECD Publishing, Paris, <https://doi.org/10.1787/5a700c4b-en>. [18]
- OECD (2020), *OECD Regions and Cities at a Glance 2020*, OECD Publishing, Paris, <https://doi.org/10.1787/959d5ba0-en>. [19]

- OECD (2020), *Preparing for the Future of Work in Canada*, OECD Reviews on Local Job Creation, OECD Publishing, Paris, <https://doi.org/10.1787/05c1b185-en>. [1]
- OECD (2020), *Preparing the Basque Country, Spain for the Future of Work*, OECD Reviews on Local Job Creation, OECD Publishing, Paris, <https://doi.org/10.1787/86616269-en>. [2]
- OECD (2019), *OECD Regional Outlook 2019: Leveraging Megatrends for Cities and Rural Areas*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264312838-en>. [3]
- Pouliakas, K. (2021), *Artificial intelligence and job automation : an EU analysis using online job vacancy data*, Publications Office of the European Union., <https://doi.org/10.2801/305373>. [15]
- Taska, J. (2020), "Concentration in US labor markets: Evidence from online vacancy data", *Labour Economics*, Vol. 66. [20]
- Tsvetkova, A. et al. (forthcoming), *Representativeness of Lightcast web-scaped vacancy data. An assessment for largest English-speaking countries*. [16]
- Tsvetkova, A., S. Grabner and W. Vermeulen (2020), "Labour demand weakening during the COVID-19 pandemic in US cities: Stylised facts and factors related to regional resilience", *OECD Local Economic and Employment Development (LEED) Papers*, No. 2020/06, OECD Publishing, Paris, <https://doi.org/10.1787/700d91ba-en>. [12]

Annex A. National sources and matching data dimensions.

Austria

The Austrian PES has monthly and regional data regarding occupations and sectors. However, the occupational classification doesn't map one-to-one one-digit ISCO classification in the Lightcast data. The occupations database consists of 10 categories, from zero to nine, that are not consistent with the ISCO categories used in Lightcast.

Table A A.1. Austrian occupation classification

First digit of occupation group	Occupation definition
0	Agriculture and forest
1,2,3	Industry, trade
4	Trade, transport
5	Services
6	Technical Jobs
7	Administration, office
8	Health, teaching profession
9	Indefinite

Source: OECD elaboration based on Austrian PES agency, AMS.

In Table A A.2 a mapping between occupational categories from Austrian data and Lightcast is presented. Sector 9 in the Austrian dataset was eliminated since there is not a correspondent category in Lightcast data.

Table A A.2 Occupation classification for Austria

First digit of occupation group	Lightcast
4 and 5	Clerical support workers (4), Service and sales workers (5)
6	Technicians and associate professionals (3)
0	Elementary occupations (9)
1,2,3, 8	Professionals (2), Craft and related trade workers (7), Plant and machine operators and assemblers (8)

Source: authors' elaborations

Belgium

Belgian data is structured by sector and on a yearly basis, from 2018 to 2021. The data also encompasses Belgium's three regions (NUTS 1): Brussels, Flanders and Wallonia.

Germany

The data from Germany was gathered from the German PES. Data is available monthly by sector and NUTS3 regions. It includes all vacancies that have been reported to the labour agencies and employers.

The Netherlands

The PES provides a weekly update of the number vacancies by postcode and four digit ISCO occupations. Data retrieved from <https://data.overheid.nl/dataset/uwv-open-match-data>.

Portugal

The data from Portugal was extracted from the National Institute of Statistics (*Instituto Nacional de Estatística*) has a yearly, quarterly and sector arrangement. The data encompasses from the second quarter of 2020 to the second of 2021 and despite having the same sectorial classification than Lightcast data, the sector are aggregated differently. The classification can be observed in Table A A.3.

Table A A.3. Portugal's sector classification

Sector group descriptions	Corresponding NACE letter
Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply; water supply, sewerage, waste management and remediation activities	B, C, D, E
Construction	F
Wholesale and retail trade; repair of motor vehicles and motorcycles; transportation and storage; accommodation and food service activities	G, H
Information and communication	J
Financial and insurance activities	K
Real estate activities	L
Professional, scientific and technical activities; administrative and support service activities	M
Public administration and defence; compulsory social security; education; human health and social work activities	O, P, Q
Arts, entertainment, recreation and other service activities	R

Source: Portugal National Statistics

Office: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0006904&contexto=bd&selTab=tab2

Spain

The data source is the National Institute of Statistics (*Instituto Nacional de Estadística*) and it contains information related solely to sectors. Data is retrieved from Available at <https://www.ine.es/jaxiT3/Tabla.htm?t=6048&L=0>. The data is structured by quarters and encompasses sectors B to S from the second quarter of 2018 to the second quarter of 2021. Furthermore, the National Institute of Statistics defines a job vacancy as a newly created or unfilled paid position, or that is about to be vacated for which the employer is actively finding a suitable candidate from outside the company in question and intends to cover it immediately or within a specified period.

Sweden

The data for Sweden is provided by the Swedish Public Employment Service and contains monthly data from 2020 to 2022 at a regional level. Data is retrieved from https://statistik.arbetsformedlingen.se/extensions/Manadsstatistik_sid1/Manadsstatistik_sid1.html. The data contain the total number of vacancies for each of the 21 NUTS2 regions.