LIRNEasia
Pro-poor. Pro-market.

Use of online job portal data in research and in practice: A review

Merl Chandana & Vihanga Jayawickrama

July 2022

Compiled by LIRNEasia for the Asian Development Bank under project number *TA-6536 REG: Nowcasting and Disasters: Impact-Based Forecasting and Socioeconomic Monitoring (54113-001)*

About LIRNEasia

LIRNEasia is a regional digital policy think tank, founded as a not-for-profit company in 2004, working across the Asia Pacific. We conduct in-depth, policy-relevant research on infrastructure industries including the digital sector. As such, our work often extends to areas such as labor, education, agriculture, and disability. LIRNEasia has engaged in issues related to digital technology and the future of work, studying platform work across Asia using large scale surveys and deep ethnographic research. LIRNEasia is also currently using natural language processing (NLP) techniques to understand the demand for skills in the Sri Lankan job market by analyzing job advertisements in one of the country's largest online job search engines. More details of our research related to future of work can be found at https://lirneasia.net/futureofwork .

Table of Contents

Introduction

The traditional tools of understanding job markets include labor force and skills surveys, qualitative studies, and the manual analysis of job vacancies. These methods have pros and cons. What one gains in comprehensive coverage (e.g., from a nationally representative labor force or skills survey), one compromises in resources (due to cost and time required). Many methods often fall short in capturing the complexity, the variability, and the pace of change of labor markets at a level of granularity that is useful for policy makers and other seekers of such information. It is in this context that OJPs have emerged as a promising data source that can bridge information gap.

OJPs, in the ideal case, offer near real-time information on the current skills demanded by employers. They also enable comparisons across time and a wide range of occupations, industries, and geographies. They also allow for the early detection of emerging labor market trends, providing job seekers, employers, and policymakers with a forward-looking analytical tool. Further, job portals with advanced functionality often contain data about the job seekers and their behavior on the platforms, which can provide additional insights into the supply of skills, preferences of job seekers and job searching patterns on these job portals.

However, there are several challenges that need to be overcome before OJPs can be used to draw reliable conclusions on labor markets. The principal concern with using OJP data for economic analysis is that their data are not representative of the full job market. Not all vacancies are advertised, and even among the advertised, there are differences in coverage from one OJP to another based on their target market segment, language, approach used to collect ads, etc. The second important challenge is data quality. Given that the data generated on OJPs aren't collected with research objectives in mind, there are no common standards for vacancy formats, schemas, and job classifications. Even within the same country, there can be significant differences in the nature, the amount, and the quality of data captured by different job portals. And finally, the legal and ethical frameworks for utilizing job portal data are not always clearly established. This concern is even more pressing when data about individuals and their behavior on job portals are used for analysis. In the Global South, these challenges are amplified by low levels of digitization, low levels of digital skills, and high levels of informality in the labor market, among other challenges. This review takes a detailed look at the challenges of using OJP data for labor market analysis.

Due to the recency of this field and differences in country contexts and technical features of job portals, there are no standardized ways of conducting job portal analysis. Researchers studying OJPs employ a variety of techniques drawing from different disciplines including, statistics, econometrics, and computer science. A thorough understanding of methods and techniques available along with their strengths and limitations can enable the selection of the right combination of techniques for a given research question.

This review is arranged around the following key themes:

1. How has online job portal data been used for labor market analysis?
2. Examples OJP usage in real world applications.
3. Limitations and challenges of using OJPs and existing ways of addressing them.
4. Other data sources that complement OJP data.
5. Processing steps, methods, and techniques used in collecting and processing OJP data prior to analysis.

How has online job portal data been used for labor market analysis?

The information captured by online job portals (OJPs) can be broadly categorized into two types based on their information content and the focus of past work.

1. Online job vacancy data (OJVs)
2. Non-vacancy data

For each of the two types of data, this section examines the variables captured, explores the kinds of research questions that can be answered and provides an overview of studies that have been conducted.

Online job vacancy data (OJV)

An online vacancy is an advertisement for an occupation which can be broadly understood as an employer looking for a set of skills to perform one or more tasks.

Variables captured in online job vacancy data

Significant differences exist in the content of OJVs across web platforms, between and within countries. Nevertheless, several common variables can be identified from most available sources (Cedefop, 2019).

- **occupations**: the primary piece of information in a vacancy which includes the job title and the description. Upon processing, it is possible to map the occupation to a standard job and skill classification framework such as the International Standard Classification of Occupations (ISCO).
- **countries and regions**: most vacancies include some indication of the place of work which facilitates geographical breakdowns. However, in order to make reliable inferences for a given region, the nature of the job, the type of industry, and overall geographical mobility of labor should be considered.
- **skills**: usually the most important piece of information in an OJV that is of interest to labor market practitioners. It includes employers' requirements for jobs, and skills. It should be noted that sometimes the vacancy content is not an accurate reflection of the skills needed for the jobs since some talent-seekers might include a wish list of skills to find the best talent regardless of the job requirement, while some others might only use vacancies as a primary filter for shortlisting candidates.
- **time dimension**: most vacancies contain some information about the date the vacancy was announced and the duration it was open, allowing monitoring and comparisons in the labor market.
- **other variables**: vacancies may also contain a combination of other variables, such as the wage, contract type, non-skill requirements, working conditions or required experience which allow for the slicing of data in several ways.

The kinds of questions that can be answered with vacancy data

OJV analysis is usually driven by two main objectives: to monitor labor markets across various dimensions and to understand employers' skill requirements. More specifically, OJVs have the potential to add more clarity to questions such as:

- For which occupations is demand increasing most? In which sectors or regions?
- What combination of skills are sought by employers in these 'top jobs? What new types of jobs are emerging?

- What are employers' demands for specific skills in specific jobs? How does this differ across countries, regions, or sectors? What new skills are employers demanding? In which jobs?
- Considering the set of skills required in different jobs, what possible career moves are there for jobseekers? Which jobs, although different, require a similar set of skills?
- Apart from skills, what other requirements appear to be most salient? What groups of people might be at a disadvantage, due to certain requirements?

Past work leveraging OJV data

Skills analysis

The primary role of OJPs from a labor market perspective is to bridge the gap between the demand and supply for skills. Therefore, OJVs have been extensively used for studying various aspects of skills including soft skills, technical skills, industry-specific skills and transferable skills. While most soft skills are required for jobs spanning a wide range of industries, the skills that are deemed as most important could vary with the industry. Technical skills, with some exceptions, are often specific to the industry.

An analysis of employability skills (as defined by the Department of Education, Science, and Training) in job vacancies was conducted in Australia via an integrative review of 40 previous studies (Messum et al., 2016). The authors found communication and teamwork skills to be far more pronounced than job-specific skills across most jobs. In another study, variations of skill requirements across the dimensions of time, the scale of the company and salary was conducted via a combination of multi-criteria analysis and skill popularity-based topic modeling (Xu et al., 2017). Results identify mobile development skills to be more salary-oriented and data-driven skills to be steadily gaining popularity. More recently, the use of Artificial intelligence (AI) techniques for skills analysis has become quite popular. A recent European study leveraged AI and graph-theory methods to map out occupation/skill relevance and similarity over the European Labor Market (Giabelli et al., 2020).

The identification of sector/industry specific skills is another popular use-case of OJV analysis. This (Messum et al., 2011) discusses essential skills for health managers presented in both online and offline vacancy advertisements in New South Wales, Australia. Furthermore, soft and hard skills that are specifically in demand in the IT industry have been analyzed by Ternikov (2022). Here, skills requirements are analyzed at the industry level as well as at the level of job clusters.

Other applications of skills analysis include improving the accuracy of job recommendations, adjusting the curricula at skill-developing organizations, assisting in talent search procedures, exploring skill salience, and predicting the demand for skills. A detailed survey on skill identification using job ads was done by Khaouja et al. (2021) and it analyzes recent trends, evaluates methodologies, and provides an in-depth comparative survey of research papers related to skill identification from online job ads. Further, a recent review by Fabo & Mýtna Kureková (2022) notes that online labor market data has also been used to study skills within the contexts of school-to-work transition, new occupations, and lifelong learning.

Labor market monitoring and analysis

Given that vacancy data is a representation of the demand for skills, it can be used to keep track of various dynamics in the labor market. Changes in the labor market over time, geographical location, and demographic features, as well as the effect of shocks to the economy, can be analyzed using job vacancy data. Given their longitudinal nature and near real-time access, Fabo & Mýtna Kureková (2022) note that vacancy data has been a popular data source for studying labor market fluctuations.

Acemoglu et al. (2020) studied the impact of AI on labor markets, using establishment level data on vacancies with detailed occupational information comprising the near universe of online vacancies in the US from 2010 onwards. While the study discovered AI-related job vacancies to be growing fast during the period in consideration, no significant effect of AI exposure on employment or wage growth was been found. In addition, Azar et al. (2020) explores the concentration of labor market in the US across occupations and geographical regions, emphasizing the high correlation between higher market level concentrations and lower wages in the labor market as opposed to the product market.

More recently, a study published by the OECD (2021) dives into the details of the effect the COVID-19 pandemic had on the labor market. The study focuses on the changes to the total volume of OJVs as well as changes specific to certain sectors and working arrangements. According to the findings, though the total availability of OJVs have taken a significant fall during the pandemic, the demand for certain skills in healthcare such as “intensive care” and “basic patient care” have been increased. They were also able to notice the variance of these trends between countries, sectoral differences in fluctuations, and the change in skill requirements. Furthermore, the quantity of OJVs offering remote working arrangements has increased as well.

Testing of sociological/economic theories

OJV data has also been used by economists, sociologists and other labor market practitioners to test and validate hypotheses and theories such as gender discrimination, and unemployment. Gender discrimination in online job advertisements in China is explored in Kuhn & Shen (2013). Findings reveal that statements regarding the preferred gender are more often seen in jobs that require lower skill levels, and that they follow the preference the associated firms have regarding job-gender matches rather than a logical model. An interesting correlation between wage levels and soft skills associated with genders is presented in Calanca et al. (2019), with skills that are often viewed as ‘female’ skills being associated with wage penalties. The research also shows that soft skills listed in vacancies could assist in predicting the gender composition of professions.

A study by Card et al. (2021) suggests that the elimination of gender preferences from job advertisements could result in an increase in gender diversity of hired employees in jobs that are typically targeted towards a certain gender. The research was conducted on the Austrian job market, where the specification of gender preferences in job advertisements has been illegal since 2005. A thorough examination of the relationship between wages and unemployment is presented in Faryna et al. (2022). Variations in the relationship across different geographical regions and skill segments are presented in the study as well.

Other examples of studies testing economic and sociological theories using OJP data include: the value of the migration experience in employers’ demands (Kureková and Žilinčíková, 2018); the relationship between firm credit crunch and employee job search behavior (Gortmaker, Jeffers, and Lee 2021); the role of occupational mismatch in explaining the productivity puzzle (Turrell et al. 2021); and links between the introduction of unemployment benefits, job searches and job postings during the recent recession in the US (Marinescu, 2017)

Non-vacancy data

Apart from the OJVs that are advertised by employers, OJPs contain a wealth of information about registered jobseekers and user activity on these job portals. While research and already implemented applications of this data is limited compared to vacancy data, it has the potential to unlock unique and diverse insights into labor market dynamics (Fabo & Mýtna Kureková, 2022).

Variables captured in non-vacancy data

The three main categories of non-vacancy data are jobseeker data, employer data and transaction (website activity/traffic) data.

1. **Jobseeker data**: usually found in the forms of user profiles and CVs (Curriculum Vitae), jobseeker data includes biographical details, educational qualifications, skills, and past work experience of users of OJPs.
2. **Employer data**: details about companies which have a presence in OJPs found in the form of company profiles. Some information about the employer advertising for the given vacancy can generally be found within the vacancy itself.
3. **Transaction data**: As jobseekers, employers and other visitors navigate through OJPs their activity generates a trail of data points that can be made to draw diverse inferences. These inferences are quite rich if they belong to registered users who use a log-in to access the job portals, since it allows for longitudinal analysis. Some of the website visit based information includes,
   a. Number of clicks
   b. Number of received applications per vacancy
   c. Number and the timing of applications submitted by a job seeker
   d. Time spent on site/page
   e. Click streams that captures website navigation
   f. Revisits

The kinds of questions that can be answered with non-vacancy data

Non-vacancy data can be used both on its own, and in combination with vacancy data to answer and add clarity to a variety of questions. Some of them might include:

- How many vacancies were viewed by the average job seeker before an application was submitted? Does it vary according to industry, geography, or the time of year?
- What skills did the job seeker profiles with the greatest number of views and longest times spent on profiles have in common?
- What vacancies received the most vs least number of views and applications? How fast are they received?
- Are there differences in job searching behavior among young vs more-experienced jobseekers in the labor market?
- For vacancies with identical job requirements published around the same time, was there a significant difference between the number of views/applications?
- Do jobseekers belonging to certain groups (age, experience-level, ethnicity, gender) show differences in their job search and application behavior?
- Which vacancies are most suitable for a given applicant, and which applicants are most suitable for a given vacancy?

Past work leveraging non-vacancy data

Job seeker data combined with job portal activity presents a plethora of opportunities to understand and reduce labor market inefficiencies. Adrjan & Lyndon (2019) developed a new measure of labor market tightness using the number of clicks on a job posting shows that advertised salaries are significantly higher in jobs where the supply of potential workers is low relative to demand. Brenčič (2014) found that employers and job searchers prefer to visit job portals with more postings. However, once on the site, the number of postings that a typical visitor reviews are not affected by the number of available postings. A study conducted by Hensvik et al., (2020) on Sweden's largest online job portal found that Job seekers decreased their search intensity by 15% during the 3 months after the COVID-19 outbreak and vacancy-level tightness increased by 25% during the same time period.

Another category of research conducted using non-vacancy data on job portals is experimental studies. These kinds of studies for economic analysis using online job portal data are relatively novel, but the experimental form of these studies is quite popular among software engineers (who use it to test new features on websites) making them relatively straight forward to conduct. In a randomized controlled trial conducted in an online job portal in India, Yamauchi et al., (2018) examined the impact of noncognitive (socio-emotional) skills on job market outcomes. Results of the study seem to indicate that employers exhibit a higher interest in candidates when they have access to knowledge about personality data of the candidates.

Analysis of transactions and portal activity to draw inferences and provide recommendations are also emerging as a prominent area of research. Matsuda et al., (2019) analyzed a Pakistani job portal and found out that jobseekers who are male, young, and with higher current salaries and educational degrees submit more applications. They also found out that many jobseekers applied for jobs that offer less than their desired salary level. Lu et al., (2013) developed a hybrid recommender system exploiting the job and user profiles and the actions undertaken by users in order to generate personalized recommendations of candidates and jobs

Examples OJP usage in real world applications

In addition to academic studies which have used OJPs to study labor market there have been several real-world applications of OJPs used by decision makers in the labor market. Nitschke et al. (2021), and Fabo and Mýtna Kureková (2022) highlight several such applications:

1. The National Skills Commission of Australia uses big labor market data (including OJPs) in their Jobs and Data Infrastructure (JEDI) tool (National Skills Commission, n.d.). It uses a mix of traditional sources of data and job posting data to identify 25 emerging occupations within the Australian labor market that are not well articulated in standard occupational taxonomies.
2. In Singapore, the government has launched the SkillsFuture project which uses the data from job postings, combined with insights from stakeholders' interviews, to support policy and program design (Job-Skills Insights, 2022)
3. The European Union (EU) partners with labor analytics company Emsi Burning Glass (EBG) to develop CEDEFOP’s (European Centre for the Development of Vocational Training) online job vacancy analysis system (CEDEFOP, n.d.). Cedefop, an EU agency focused on the development of European VET (Vocational Education and Training) policies, has been working with big labor market data since 2015.
4. The World Bank and the government of Malaysia released a report on Malaysia’s skill shortages and critical occupations (World Bank, 2019). They used postings data to learn the skill and experience requirements of high-demand occupations and help create their list of critical occupations. This list is then used to align workforce development policies with employer demands
5. As part of the O*Net project in the United States, online job vacancies were used to determine “hot technologies” on the basis of employers' job postings (Lewis & Norton, 2016)
6. Indonesia created a critical occupation list to highlight shortages and potentially strategic investment areas (World Bank, 2018a). Additionally, they have an Online Skills and Vacancy Outlook initiative that collects online job postings by occupation. These two initiatives work to analyze skill imbalances and help policymakers to make investments in training programs and adjust incentives.
7. In New Zealand, Tokona Te Raki, an indigenos social innovation lab, is using big labor market data to drive longer-term systemic change to boost Māori success and tackle inequality (Tokona te Raki. 2020). They aim to produce an understanding of current labor market data to support better employment outcomes for Māori and inform the

business case for further investment in tools to enable future indigenous workforce development.

8. In the United States, OJP analysis has led to Colorado's new Pay Transparency Law which requires employers to (1) post compensation and benefits information for each job posting for Colorado jobs and (2) internally post promotional opportunities to current Colorado employees on the same day and sufficiently in advance of promotion decisions (Muhleisen et al., 2021). Changes in wages based on these policy changes show up immediately in online job posting data, which enable the changes in wages or other job characteristics to be tracked.

Limitations and challenges of using OJPs and existing ways of addressing them

While OJPs have emerged as an important source of labor market information, they come with several limitations which need to be considered before they are used to explore labor market questions. Given the disproportionate amount of analysis done using vacancy data compared to non-vacancy data, the problems identified, and the solutions proposed have mostly been centered on vacancy data. However, they can be adapted and applied for both vacancy and non-vacancy data since methodological issues arising from their use have common roots.

Representativity

The universe of jobs advertised online is not equal to the universe of new jobs that exist in a given labor market. This can occur due to a variety of reasons

Issues

1. Penetration of OJPs: OJPs capture only part of the demand & supply dynamics of the labor market. Low internet penetration and lack of basic digital skills across the population are the key parameters influencing employers' decisions on the extent of use of OJV portals as a recruitment channel. Therefore, certain sectors and types of occupations are overrepresented in OJPs. Further, more recently, dedicated professional networking sites such as LinkedIn and other social media channels such as Facebook and Twitter have also emerged as alternatives to OJPs.
2. Different hiring practices of different industries and geographies. Even among countries with high internet-penetration, the number of jobs that appear can differ significantly. For example, Van Loo and Pouliakis (2020) reviewed online job markets in the EU28 countries during 2019, and concluded that in countries such as Estonia, Sweden and Finland, the proportion of vacancies published online approached 100 per cent, while in others such as Denmark it accounted for around 50 per cent.
3. OJVs represent only part of the job demand - not all job vacancies are advertised on-line (some are filled in house or through professional networks without ever advertising), and some jobs are more likely to be advertised on-line than others. It can be expected that data are subject to occupational or qualification bias. For example, according to a 2014 study by Carnevale et al. on the US Labor Market using Burning Glass Technologies (BGT) data:
    - Online job ads data overrepresent job openings for college graduates
    - Between 60 and 70 percent of jobs are posted online
    - More than 80 percent of jobs for those with bachelor's degrees or better are posted online
    - Job ads overrepresent industries that demand high-skilled workers
    - White-collar office and STEM occupations account for the majority of ads

Potential solutions

Based on their review of theoretical and empirical studies, Fabo and Mýtna Kureková (2022) distinguish between inductive and deductive approaches to analyzing labor market data. The inductive approach is bottom-up and exploratory, often concerned with understanding the underlying qualities of labor markets (e.g., studying skills and trends using longitudinal data). The deductive approach concerns itself with theory-testing based on probabilistic and inferential statistics methods. As such, the representativeness and generalizability of the data is less of a concern for studies based on the inductive approach as opposed to those based on the deductive approach. However, the authors note that most studies (both inductive and deductive) do not discuss any aspect of bias of their data: they are not concerned with broader generalizability beyond briefly acknowledging the issue in footnotes or as a short remark

In instances where some attempt has been made to account for the bias in data, some of the earliest techniques including sectoral and occupational structure of the Labor Force Survey (LFS) to adjust the estimates of the coverage of online vacancies. However, Kurekova et al. (2015) who reviewed different methods of addressing representativity of OJVs concluded that given the LFS is not a reliable measure of the structure of the demand side of the labor market. LFS captures the existing stock of jobs within a labor market, whereas vacancies represent new openings within it. However, other solutions have since been proposed by other researchers to mitigate and address some of the issues of representativity in OJV data.

1. In using data from OJPs, limit the focus on the segments of the labor markets where the coverage bias is less likely to be an issue. Examples include focusing on graduates' CVs and vacancies targeting this labor market segment or on sectors and professions that are, by their nature, characterized by widespread access to the Internet (e.g., IT and Software jobs).
2. The diversification of data sources and approaches has been used in many studies. In addition to online job ads, other types of data sources can be analyzed in parallel, (and findings can be verified/supplemented/modified) such as practitioner literature, administrative data, or interviews of sectoral HR professionals.
3. Market coverage and technical advancement of the online job portal(s) in a given country need to be assessed in each country. In countries where a dominant portal exists, collected vacancies might be the best available source. These alternatives need to be weighed against the costs of collecting data by other means. Using job advertisements from an established portal and interpreting the results with caution to avoid potential biases can be a valid and acceptable choice.
4. Statistical approaches. According to Fabo and Mýtna Kureková (2022), three types of statistical approaches have been used to account for the non-representativity of online labor market. These include:
   i. Rule-based and statistical techniques such outlier approaches and de-noising data (covered in the next section on data quality in this report)
   ii. Using dedicated surveys on vacancies that help account for the bias through proper weighting. For example, Turrell et al. (2019) used data from a leading online job portal (reed.co.uk) with a vacancy survey by the Office of National Statistics, UK to compare the mean annual ratios of different sectors and found no significant biases in professional and scientific activities, ICT and administration, whereas public administration and manufacturing appeared to be mostly absent from the online portal
   iii. Using the sheer size of the online data has been exploited as a strategy or a justification for not making additional adjustments. It has been argued that the richer the data, the better models can be developed, where the sheer number of observations typically present in big data acts as a "self-corrective" mechanism (Mezzanzanica & Mercorio, 2019)

Varying quality of information sources

Given that the data found on OJPs are not collected with research objectives in mind, it is perhaps not surprising that they cannot be directly used to derive labor market insights. There are several data quality issues that need to be overcome before OJP data can be used to generate reliable labor market insights (Cedefop, 2019).

Issues

1. In most countries, the OJP market comprises multiple actors with different business models. There is usually no single source of all online job vacancies; the volume, the variety and quality of the data depend on the portals selected for analysis.
2. Reliability of vacancy information & self-reported information of jobseekers are also difficult to assess. For example, Internet job boards can be flooded by resumés that in fact no longer correspond to people who are searching for a job – known as “stale” resumés (Fabo & Mýtna Kureková, 2022).
3. To compare job titles, skills and competencies in a standardized manner, ontologies need to be developed to sort and organize a diverse and complex universe. Despite enormous effort in developing them, they are still imperfect and full of simplifying assumptions that go into them; therefore, they are likely to contain systemic errors that can only be corrected over time.
4. In some instances, vacancies are published on several web sites while in other cases some vacancies do not necessarily correspond to an actual job opening (ghost vacancies). Further, skills listed in a vacancy notice do not reflect the full job profile; employers tend to list only critical skills and qualifications to ‘filter’ job applicants. When there are multiple fields of information on an online job vacancy the accuracy of extracted labor market information can vary across data fields.
5. Vacancies and the information they contain need to be processed, often in several steps, to produce viable data. Sometimes data might not be in direct machine-readable form. For example, in Sri Lanka, vacancies on the most popular OJP, topjobs.lk are posted in image form which require optical character recognition (OCR) to translate into textual format. Even the most up-to-date techniques of OCR are still imperfect.

Potential solutions

While there are no perfect solutions to remove data quality issues altogether, explicitly recognizing all data quality issues upfront will help rectify some of the issues and inform the limitations of the inferences that can be made with a given dataset. Some of the specific ways of addressing data quality limitations are covered later, under data pre-processing techniques. However, it often helps to manually inspect the given dataset based on a predetermined strategy, to uncover the issues that might be prevalent in each dataset. Additionally at a high-level,

1. Verification of the quality of data with the data provider, wherever possible, is key. The objective is to identify data errors and anomalous patterns; this is not to measure overall data quality, but to remove data that may not be reliable, understand the roots of the data quality issues and generate solutions on addressing some of them.
2. Working towards a set of best practices to solve common problems by documenting the processes, suggestions, and revisions; ontologies and training sets can be published under creative common licenses to enable potential contributors to evaluate and improve them.
3. Using open-source tools for many of the computational tasks of data quality rectification. Some of the examples include the Tesseract framework for optical character recognition (OCR) from images, the natural language toolkit (NTLK) for text

processing, and Genism for information extraction through topic modeling. These tools are considered to industry-standard with an active community of software developers making constant improvements. They also allow methods and research developed using OJP to be comparable and reproducible.

Compliance and privacy concerns

Given OJPs contain data that was not meant for analytics purposes, analyzing them raises a number of ethical questions, especially around compliance and privacy. However, as is the case with representativity, many studies do not address them explicitly, beyond briefly acknowledging the issue in footnotes or as a short remark. There are two major issues in this area.

Issues

- Web scraping, which is one of the primary methods of collecting gathering OJP data, involves querying a website repeatedly and accessing a potentially large number of pages. For each of these pages, a request will be sent to the web server that is hosting the site, and the server will have to process the request and send a response back to the computer from which the scraping operation is run from. This raises several ethical concerns,
    i. Too many requests sent during a short amount of time consume server resources that can prevent other "normal" users from accessing the site during that time, or in a worst-case scenario even cause the server to crash.
    ii. In certain circumstances web scraping can be illegal. Sometimes the terms and conditions of the web site being scraped specifically prohibit downloading and copying its content
- Most OJPs contain job seeker profiles and can capture their activity on the job portal across different interactions. While this data can inform better public policy, improve job seeker experience on the portals, and provide employers with better matches, OJPs contain several personally identifiable attributes registered individuals, that gives rise to serious privacy concerns.

Potential solutions

- There aren't any global frameworks on the legality and ethics of web scraping, and this is an area which can be full of seemingly contradictory opinions. While some argue that web scraping can result in serious ethical controversies if conducted without care (Krotov et al., 2020), some others argue that web scraping is generally a tolerated practice, provided reasonable care is taken not to disrupt the "regular" use of a web site and the data is publicly available (the content that is being scraped is not behind a password-protected authentication system). If fact, it is often argued that web scraping is no different than using a web browser to visit a web page, in that it amounts to using computer software (a browser vs a scraper) to access data that is publicly available on the web (Monash Data Fluency, 2022). However, the best practice in obtaining OJP data is to inform OJP operators about the intended data collection and, where possible, come into written agreements for data sharing and usage. In some cases, portal owners can even grant direct access to data via API, which significantly reduced the amount of data pre-processing needed.
- Just like web scraping, the legal and ethical frameworks for obtaining and using such personally attributable job portal data are not always clearly established. Most portals with registered jobseekers usually have their consent to an agreement containing terms and conditions surrounding data usage. However, in practice, it cannot be assumed that all jobseekers fully understand what they are giving consent to. Countries which have personal data-protection laws are likely to have provisions

which govern the use of personal data on OJPs. Where such laws and frameworks do not apply, it is always best practice to aim for informed consent.

Other data sources that complement OJP data

As discussed in the two previous sections, even though OJPs can provide rich and diverse insights, they still represent only an incomplete picture of the labor market. Additional sources of data are needed to address the limitations of OJPs, answer other questions about the labor market, and validate findings obtained from OJPs. By recognizing the attributes, relative strengths and limitations of the diverse types of sources available, the findings of any labor market insights can be expanded and made more robust (Nitschke et al., 2021). These complementary data sources can be divided into two broad categories: (I) traditional sources of labor market data and (II) novel sources of labor market data

Existing/traditional sources of labor market data

Labor Force Surveys

Across the world, the primary source of information on labor market dynamics is government collected data, often collected via labor force surveys (LFS). This data is usually maintained by individual government organizations as well as some international organizations. The primary objective of these surveys is to collect information on the volume of people working in different occupations. Labor Force Surveys are typically designed to answer research questions like: What are the jobs in my labor market? Which occupations have seen decline or have increased in the long term?

Given the robust statistical sampling methods used during the design of these surveys, LFS data are representative with stable definitions and taxonomies adopted to ensure that results are comparable across time and compatible with similar surveys across countries. Further, aggregate data on labor market outcomes, including employment by location, industry and occupation are widely available for free to the public online. However, it is important to understand the ways in which LFS data is limited for the purpose of understanding labor markets. As noted in the previous section, LFS captures the existing stock of jobs within a labor market, whereas vacancies represent new openings within it. Therefore, it is not advisable to use LFS data to weight online vacancy data as a means of correction. However, LFS data often gives the most reliable picture of the distribution of jobs that will help deepen the contextual understanding of a given labor market.

Further LFS data also suffers from several other limitations. They lack the granularity to understand variables such as skills at the occupation level, education, experience, and other key labor market dimensions. Further the frequency with which data is collected and made available is limited by the labor-intensive method of survey data collection. In general, due to prioritizing time series analysis and the ability to maintain consistency, government taxonomies are updated infrequently. This means that even within existing jobs emerging trends in labor markets are difficult to capture using LFS data.

Skills, knowledge, abilities, and competency data

In understanding labor markets, it is useful to leverage a common taxonomy of skills, competencies, and knowledge domains to facilitate consistent comparisons across industries, geographies, and time. Therefore, skills knowledge and competency datasets are helpful for determining which skills are important to different occupations and understanding how occupations' skill requirements compare to each other. They can address research questions such as: What skills are required for this occupation? How do occupations compare to each other in terms of skill requirements?

Altamirano and Amaral (2020) note that O*NET and ESCO are the two most well-known skill taxonomies. These taxonomies classify occupations and skills and map them to each other. Both ESCO and O*NET offer deep levels of granularity, with thousands of variables and unique skills collected for each occupation. This makes them extremely valuable for understanding in-depth the competencies, skills and abilities required for various occupations. Both O*NET and ESCO are compatible with their respective government occupation taxonomies (and in the case of ESCO with the ISCO taxonomy).

However, these standardized jobs and skills taxonomies are limited in several important ways. Given the extensive process of employee and expert consultations that underlie their methodology, skills and job taxonomies are quite slow to update. The slow update process makes it difficult to track how occupations change over time in terms of the skills they require, something that is becoming even more salient as digitalization transforms occupations. A second important limitation of this data is its limited relevance across economies. O*NET is based on the American economy, whereas ESCO is based on the European Union economy. Differences in industry cultures could affect the types of skills deemed important for a job, so these frameworks are limited in the economies they can be applied to.

Technology adoption and other primary survey data

Another source of important labor market information are surveys commissioned by governments and private sector organizations. When new technologies (e.g., Artificial Intelligence) are perceived to be disruptive to labor markets, technology adoption survey data seeks to measure emerging technology adoption across companies and locations. Stakeholders use this data to understand how to prepare for changes in specific occupations and the labor market in general.

Further, there can be dedicated surveys which aim to capture specific aspects of the labor market. For example, the Office of National Statistics in the UK conducts “The Vacancy Survey”, which is a regular survey of businesses that provides an accurate and comprehensive measure of the total number of vacancies across the economy (Vacancy Survey, n.d.). Therefore, this survey presents a number of opportunities to complement the analysis of OJPs, including the weighting of vacancies to address the lack of representativity (Turrell et al., 2019).

The quality of the information that can be obtained from these surveys is often determined by their survey designs. While government commissioned surveys might be representative of the studied population, they might not reveal detailed insights into specific critical issues. Conversely, private sector surveys might focus on a select group of people for deeper insights instead of developing a representative sample. Further, given their one-off nature, these types of ad-hoc surveys rarely generate insights that are comparable across time and geography.

Other sources

Apart from the sources mentioned above, there are other traditional sources of data that might complement the information derivation process of a given labor market. There are professional bodies and sector organizations (e.g., institutions for accountants, construction sector professionals) that have stated objectives of monitoring and responding to demand/supply dynamics of labor markets. Furthermore, there could be sector and region-specific local knowledge developed by education and training providers, career guidance practitioners and human resource experts in various sectors and industries.

Novel sources of labor market data

Online social networks for professionals

Online social networks allow professionals to connect with each other, find jobs and learn new skills. While there are several emerging and niche alternatives (e.g., AngelList, Jobcase), LinkedIn is by far the most popular online social network for professionals. LinkedIn collects data from professionals who upload their work and education histories, and from employers that advertise on their site. It claims to have more than 800 million members and more than 58 million companies (LinkedIn, 2022).

The unique characteristics of LinkedIn data provide detailed insights into work history and skills of professionals, while also unearthing the nature and the extent of professional networks and their impact on job seekers, employers, and labor markets more broadly. Online social network data for labor market analysis suffer from some of the same limitations of OJPs; they lack representativity, unstructured terminology making comparisons difficult and issues surrounding privacy of personally attributable data.

Human capital management data

Human capital management (HCM) solutions providers capture a wealth of labor market information as a by-product of business operations. Information extracted from HCM operations include employer name, industry and number of employers on payroll; employee name and demographics; job title/occupation, job type – full-time, part-time, permanent, temporary, internship, remote, etc.; payroll distribution including wages and tax payments benefits packages. ADP (Automatic Data Processing) and Kronos are two examples of third-party intermediaries providing HCM data.

As such HCM data can provide real-time insight into the employment activity of labor markets. The data can inform whether demand for talent or hiring is increasing, decreasing or stagnant by location, industry, and employer. It can serve as an early indicator of how unemployment may change or has changed over a certain period. Wage analyses can inform whether a labor market is tight or loose based on wage inflation, deflation, or stagnation. Insight into hours worked can inform whether overall economic activity is strong, waning or recovering. Compensation and benefits analyses can inform how incentives impact hiring, turnover and hours worked. These data can also highlight worker mobility, pay inequities, diversity, worker performance.

Like job postings and work history data, HCM data do not reflect the full universe of a labor market as not all businesses use third parties for HCM. Certain industries and businesses like agriculture/farming and small family-run operations are likely to be underrepresented. Self-employed individuals, contractors, and even gig workers are also likely underrepresented in these data.

Online learning platform data

Online learning platforms capture a wealth of data around how people acquire skills. Information extractable from online job platforms includes different combinations of skills and subjects' people are learning, who is learning them and where. The most and least effective ways of teaching to train people in a particular skill.

The granular data online learning platforms collect can inform the types of incentives and feedback loops that encourage and support learning. The data can reveal what people are choosing to learn, how they perform and how outcomes can be improved. Competitive analyses can be performed around the competencies being acquired across regions. Temporal analyses can be performed to determine whether certain political, economic, professional, or personal events affect when people are most likely to seek new skills and complete training.

Data gathered by online learning platforms will be limited to people and regions with internet access. It cannot inform learning styles, incentive structures or outcomes of traditional classroom learners (Mohan, et al., 2020). Datasets may be skewed toward learners from particular industries, professions and with certain levels of educational attainment.

Online gig economy data

Intermediaries that operate within online labor markets or the gig economy create unique opportunities for researchers to analyze self-contained markets and conduct natural and field experiments. Platforms collect granular data about who takes gig jobs, where workers are located, and the kinds of job workers are willing to accept and with what incentive. These platforms can provide insight into the factors that increase worker performance and reputation. These data can inform what political, economic, environmental, social, local market conditions, etc. compel workers to participate in the gig economy. It can also inform the advantages and the downsides of algorithmic control found on many gig platforms (Wood, et al., 2019).

Data from online labor markets face shortcomings that may over- or underrepresent workers from certain socioeconomic backgrounds and regions. Data from certain vendors may be more robust in metropolitan areas compared to suburban and rural areas. For example, suburban and rural areas tend to have higher rates of vehicle ownership and may rely less on rideshare and delivery services. Workers in certain regions may also be more or less likely to engage in task-related gig work depending on the opportunities of the proximate labor markets.

Processing steps, methods, and techniques used in collecting and processing OJP data prior to analysis

Upon the identification of the data needed for a given project or application, the collection of online job portal data involves several steps (Cedefop, 2019).

1. Data ingestion: obtaining the data from the source into your own systems.
2. Data Pre-processing: converting the data into machine readable form, removing the noise in the data and cleaning it of irrelevant content.
3. Organization: it translates the relevant content of the OJV into a database organized into a schema that best fits the project needs. This final database feeds various automatically updated dashboards or analysis to produce labor market and skills intelligence.

Data ingestion

Data from the selected sources can be accessed from website front ends and/or back ends (databases and systems powering the websites that the operator can provide access to). While in some cases this might involve the transfer of data through physical storage devices, it is far more common to access the data through the internet. Various techniques can be used:

- Direct access via **application programming interface** (API) allows download of vacancy content directly from OJV portal databases. This direct access requires a formal agreement from the website operator and is subject to maintenance and agreement costs. Data collected in this way is of the highest quality of the different methods and can be downloaded much faster.
- **Automated retrieval from the web (web scraping)**. In instances where data is not directly accessible through an API, scraping is a technique that is widely used in academic research and in industry. However, as has been outlined in section 4.2, the legal and ethical implications of web scraping are poorly understood (Krotov et al., 2020). However, when carried out with the data provider's explicit approval and guidance, this can be a method of obtaining data that is otherwise difficult to obtain. Web scraping can usually take two forms:

- i. **Direct Scraping** is used to extract structured data from websites. Using web scraping implies that data is already structured on the web page and can be extracted precisely by knowing the exact position of each field on the web page. As specific web scrapers must be programmed for each website, this is ideal for sites which contain many vacancies.
- ii. **Web Crawling** uses a programmed robot to browse web portals systematically and download their pages. Crawling is much more general compared to scraping and is easier to develop. However, crawlers collect much more website noise (irrelevant content), and more effort is needed to clean the data before further processing.

Data pre-processing

Data sources vary in type, quality, and content. To develop a database suitable for subsequent analysis, various pre-processing steps need to be carried out. These can include:

Optical character recognition (OCR)

The vast majority of OJP data will be a combination of structured/unstructured text and query results from a database. However, there are some job portals which post their vacancies in image form. This image data needs to be digitized into machine-readable text before computational methods can be used to conduct any labor market analysis.

Optical character recognition generally involves three main steps.

1. Pre-processing images – before using computer vision techniques for character/feature extraction, images need to preprocess to boost the chances of the characters being recognized. Some of the preprocessing techniques include fixing orientation, line removal, enhancing contrast and converting images to black and white.
2. Feature extraction – this step involves the lines and strokes to isolate characters and using pattern recognition techniques to match them to a known set of characters. This step is usually handled by OCR engines, that are commonly available through open-source programming languages such as Python and R.
3. Post-processing – even the most advanced OCR engines do not produce perfect results. Therefore, depending on the application in question, some amount of pos-processing might have to be done. OCR accuracy can be improved if the output is limited by a known set of words. For instance, this could be a set of technical words in a given field of occupation. Further, there are methods such as ‘near neighbor analysis’ that can leverage the frequencies for co-occurrence to correct mistakes. For example, ‘C++’ is likely to be more prevalent than ‘C77’.

Cleaning

OJP data, especially OJVs usually contain ‘noise’ (such as company profiles, promotions, unticked options from drop-down menus). Cleaning is a sequence of activities to identify and remove ‘noise’ from the data to improve their quality and prepare the data for subsequent analysis. While some amount of cleaning is conducted at the pre-processing stage, more advanced, goal-oriented cleaning is usually conducted at the analysis stage.

Merging

If data from more than one portal is used for analysis, there can be duplication in both vacancy data and non-vacancy data (e.g., job-seeker profiles) both within and across different OJPs. In some cases, this is a desirable property during different stages of the analysis. When this is the case, duplications of the data can be enriched by combining the information on the vacancies posted at different portals.

De-duplicating

In some cases, it is necessary to remove duplicates from the analysis. Image comparison, text comparisons, the use of meta-data (such as reference Id, page URL) can be used to identify and remove duplicates within and across different job portals.

Organization

If a given project involves the analysis or monitoring of the labor market across all sectors it is probably helpful to map the vacancies found on job portals onto a standard jobs and skills classification framework/ontology. This is not strictly necessary if a given study is limited to one or two occupations or industries. However, large scale labor market monitoring efforts such as the Pan-European Online Job Vacancies and Skills Analysis project (Cedefop, 2019) have mapped the entire universe of online vacancies onto the European Standard Classification of Occupation (ESCO). Three techniques are generally used to conduct this mapping, often in the order outlined below:

1. Direct matching of job titles with the relevant ontology
2. Using string similarity metrics to match job titles/description onto the ontology
3. Using machine learning classifiers to determine the appropriate category in the ontology

References

A4AI. (n.d.). Bhutan digital connectivity brief. Available at: https://1e8q3q16vyc81g8l3h3md6q5f5e-wpengine.netdna-ssl.com/wp-content/uploads/2021/08/Bhutan-Brief.pdf

Acemoglu, D., & Autor, D., Hazell, J., & Restrepo, P. (2020). AI and Jobs: Evidence from Online Vacancies, NBER Working Papers 28257, National Bureau of Economic Research, Inc. Available at https://ideas.repec.org/p/nbr/nberwo/28257.html

ADB. (2021a). COVID-19 and labor markets in Southeast Asia: Impacts on Indonesia, Malaysia, the Philippines, Thailand, and Viet Nam. Available at: https://www.adb.org/sites/default/files/publication/758611/COVID-19-labor-markets-southeast-asia.pdf

ADB. (2021b). Philippines' COVID-19 Employment Challenge: Labor Market Programs to the Rescue [blog entry]. https://blogs.adb.org/blog/philippines-COVID-19-employment-challenge-labor-market-programs-to-rescue

Adrjan, P., & Lydon, R. (2019). Clicks and jobs: measuring labour market tightness using online data. Central Bank of Ireland Economic Letter Series, 2019(6) https://www.centralbank.ie/docs/default-source/publications/economic-letters/vol-2019-no-6-clicks-and-jobs-measuring-labour-market-tightness-using-online-data-(adrjan-and-lydon).pdf

Afridi, F., Mahajan, K. & Sangwan, N. (2022). Employment Guaranteed? Social Protection During a Pandemic. Oxford Open Economics, 2022, 1, 1–15 https://doi.org/10.1093/oxecon/odab003

Altamirano, A. & Amaral, N. (2020). A skills taxonomy for LAC: lessons learned and a roadmap for future users, IDB technical note, no. 2072, November 2020, Inter-American Development Bank, Washington, Available at https://publications.iadb.org/en/skills-taxonomy-lac-lessons-learned-and-roadmap-future-users .

Arunatillake, N. (2021). Sri Lanka's Labour Market Amidst COVID-19: The Need for Targetted Interventions. Available at: https://www.ips.lk/talkingeconomics/2021/06/28/COVID-19-and-sri-lankas-labour-market-the-need-for-targetted-interventions/. Available at: https://www.ips.lk/talkingeconomics/2021/06/28/COVID-19-and-sri-lankas-labour-market-the-need-for-targetted-interventions/

Azar, J., Marinescu, I., Steinbaum, M., & Taska, B. (2020). Concentration in US labor markets: Evidence from online vacancy data. Labour Economics, 66, 101886 https://doi.org/10.1016/j.labeco.2020.101886

The Asia Foundation. (2021). The Impact of the COVID-19 Pandemic on Employment in Middle-order Cities of Nepal A Rapid Assessment. Available at: https://asiafoundation.org/wp-content/uploads/2021/04/Impact-of-the-COVID-19-Pandemic-on-Employment-in-Middle-order-Cities-of-Nepal.pdf

Bank of China. (2020). Recovery with Challenges -- 2021 Hong Kong Economic Outlook. Economic Review. Available at: https://www.bochk.com/dam/investment/bocecon/SY2020035(en).pdf

Brenčič, V. (2014). Search online : evidence from acquisition of information on online job boards and resume banks. Journal of economic psychology: research in economic psychology and behavioral economics. Vol. 42.2014, p. 112-125 https://doi.org/10.1016/j.jpubeco.2020.104349

Bulmer, E.R., Shrestha, A. & Marshalian, M. (2020). "Nepal Jobs Diagnostic." World Bank, Washington, DC. License: Creative Commons Attribution CC BY 3.0 IGO. Report Number: P163141. World Bank, Washington, DC. License: Creative Commons Attribution CC BY 3.0 IGO. Report Number: P163141. Available at:

https://openknowledge.worldbank.org/bitstream/handle/10986/33956/Nepal-Jobs-Diagnostic.pdf?sequence=5

Bussolo, M., Kotia, A., & Sharma, S. (2021). Workers at Risk : Panel Data Evidence on the COVID-19 Labor Market Crisis in India. Policy Research Working Paper; No. 9584. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/35292 License: CC BY 3.0 IGO.

Calanca, F., Sayfullina, L., Minkus, L., Wagner, C., & Malmi, E. (2019). Responsible team players wanted: an analysis of soft skill requirements in job advertisements. EPJ Data Science, 8(1), 1-20. https://doi.org/10.1140/epjds/s13688-019-0190-z

Carnevale, A., Jayasundera T., & Repnikov, D., (2014). "Understanding Online Job Ads Data" Georgetown University Center on Education and the Workforce. Available at https://cew.georgetown.edu/wp-content/uploads/2014/11/OCLM.Tech_.Web_.pdf

Card, D., Colella, F., & Lalive, R. (2021). Gender Preferences in Job Vacancies and Workplace Gender Diversity (No. w29350). National Bureau of Economic Research. Available at https://www.nber.org/system/files/working_papers/w29350/w29350.pdf

Cedefop (2019). Online job vacancies and skills analysis: a Cedefop pan-European approach. Luxembourg: Publications Office. http://data.europa.eu/doi/10.2801/097022

Central Bureau of Statistics Nepal. (2019). Report on the Nepal Labour Force Survey, 2017/18. Available at: https://nepalindata.com/media/resources/items/20/bNLFS-III_Final-Report.pdf

Cho, Y, & Majoka, Z. (2020). Pakistan Jobs Diagnostic : Promoting Access to Quality Jobs for All. Jobs Series; No. 20. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/33317 License: CC BY 3.0 IGO.

Danish Trade Union Development Agency. (2020). Indonesia Labor Market profile. Available at: https://www.ulandssekretariatet.dk/wp-content/uploads/2020/04/lmp_indonesia_2020_final_version-1.pdf

Deloitte. (2022). An analysis of India's labour market March 2022. Available at: https://www2.deloitte.com/content/dam/Deloitte/in/Documents/about-deloitte/in-soe-labour-market-noexp.pdf

Ernst & Young India. (2019). Developing skills in youth to succeed in the evolving South Asian economy. India country report. Available at: https://www.unicef.org/rosa/media/4496/file/India%20Country%20Report.pdf

CEDEFOP. (n.d.). European Centre for the Development of Vocational Training. Available at: https://www.cedefop.europa.eu/en

Fabo, B., Mýtna Kureková , L. 2022. Methodological issues related to the use of online labour market data, ILO Working Paper 68 (Geneva, ILO). https://www.ilo.org/global/publications/working-papers/WCMS_849357/lang--en/index.htm

Farole, T., Cho, Y., Bossavie, L. & Aterido, R. (2017). Bangladesh Jobs Diagnostic. Jobs Series; No. 9. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/28498 License: CC BY 3.0 IGO

Faryna, O., Pham, T., Talavera, O., & Tsapin, A. (2022). Wage and unemployment: Evidence from online job vacancy data. Journal of Comparative Economics, 50(1), 52-70 https://doi.org/10.1016/j.jce.2021.05.003

Fiji Trades Union Congress. (2020). Impact of COVID-19 on Employment & Business: In-Crisis Rapid Assessment: 13 May-19 June 2020 (Volume 1). Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---ilo-suva/documents/publication/wcms_754703.pdf

Giabelli, A., Malandri, L., Mercorio, F., & Mezzanzanica, M. (2020). GraphLMI: A data driven system for exploring labor market information through graph databases. Multimedia Tools and Applications, 1-30. https://doi.org/10.1007/s11042-020-09115-x

Gortmaker, J., Jeffers, J., and Lee, M. (2021). "Labor Reactions to Credit Deterioration: Evidence from LinkedIn Activity." SSRN Scholarly Paper ID 3456285. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3456285

Gounder, R. (2020). Economic Vulnerabilities and Livelihoods: Impact of COVID-19 in Fiji and Vanuatu. Oceania,Vol. 90, Suppl. 1 (2020): 107-113. https://doi.org/10.1002/ocea.5273

Government of Pakistan. (2021). Pakistan Economic Survey 2020/21. Available at: https://www.finance.gov.pk/survey/chapters_21/12-Population.pdf

GQR. (2019). Country Profile: Singapore Labor Market. https://www.gqrgm.com/country-profile-singapore/

GSMA. (2021). The Mobile Economy Asia Pacific 2021. Available at: https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/08/GSMA_ME_APAC_2021_Web_Singles.pdf

Hayashi, R. & Matsuda, N. (2020). COVID-19 Impact on Job Postings: Real-Time Assessment Using Bangladesh and Sri Lanka Online Job Portals. ADB Brief no. 135. Available at: https://www.adb.org/sites/default/files/publication/606711/COVID-19-impact-job-postings-bangladesh-sri-lanka.pdf

Hensvik, L., Le Barbanchon, T., and Rathelot, R. (2020). Job Search during the COVID-19 Crisis, CEPR Discussion Papers 14748, C.E.P.R. Discussion Papers https://doi.org/10.1016/j.jpubeco.2020.104349

ILO. (2016). Fiji Labor market update: April 2016. Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---ilo-suva/documents/publication/wcms_465248.pdf

ILO. (2017). Improving labour market outcomes in the Pacific: Policy challenges and priorities. Available at: https://www.adb.org/sites/default/files/publication/409216/improving-labour-market-outcomes-pacific.pdf

ILO. (2017). India Labour Market Update. Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---sro-new_delhi/documents/publication/wcms_568701.pdf

ILO. (2018). India Labour Migration Update 2018. Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---sro-new_delhi/documents/publication/wcms_631532.pdf

ILO. (2021). Thailand labour market update Concern remains over the drawn out impact of COVID-19. ILO Brief. Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/documents/briefingnote/wcms_829228.pdf

ITU. (2021). Digital trends in Asia and the Pacific 2021 Information and communication technology trends and developments in the Asia Pacific region, 2017-2020. Available at: https://www.unapcict.org/sites/default/files/2021-03/Digital%20Trends%20in%20Asia%20Pacific%202021.pdf

ITU. (2022). Individuals using the Internet (from any location), by gender and urban/rural location (%). https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

Job-Skills Insights. (2022). Skills Future. https://www.skillsfuture.gov.sg/skillsreport

Kapoor, R. (2022). The Mounting Challenge of Youth Unemployment in India. LSE blog. Available at: https://blogs.lse.ac.uk/southasia/2022/02/21/the-mounting-challenge-of-youth-unemployment-in-india/

Khaouja, I., Kassou, I., & Ghogho, M. (2021). A Survey on Skill Identification From Online Job Ads. IEEE Access. PP. 1-1. https://doi.org/10.1109/ACCESS.2021.3106120

Kudlyak, M., Lkhagvasuren, D., & Sysuyev, R. (2013). Systematic job search: New evidence from individual job application data.

Kuhn, P., & Shen, K. (2013). Gender discrimination in job ads: Evidence from china. The Quarterly Journal of Economics, 128(1), 287-336. http://hdl.handle.net/10.1093/qje/qjs046

Krotov, V., Johnson, L., & Silva, L. (2020). Tutorial: Legality and Ethics of Web Scraping. Communications of the Association for Information Systems, 47, pp-pp. https://doi.org/10.17705/1CAIS.0472

Kurekova, L., Beblavý, M., & Thum-Thysen, A. (2015). Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. IZA Journal of Labor Economics. 4. 10.1186/s40172-015-0034-4.

Labor Department of Hong Kong. (2021). Hong Kong: The facts, Employment. Available at: https://www.gov.hk/en/about/abouthk/factsheets/docs/employment.pdf

Lewis, P., & Norton, J. (2016). "Identification of 'Hot Technologies' within the O*NET® System." https://www.onetcenter.org/reports/Hot_Technologies.html.

Lilaiula, P. (2022). DigitalFIJI strategy aims to develop a $1bn digital economy in Fiji by 2030. https://pina.com.fj/2022/04/01/digitalfiji-strategy-aims-to-develop-a-1bn-digital-economy-in-fiji-by-2030/

LinkedIn. (2022). LinkedIn's Economic Graph: A digital representation of the global economy. LinkedIn. https://economicgraph.linkedin.com/

LIRNEasia. (2018). ICT Access and use in Nepal and the Global South. Available at: https://lirneasia.net/2018/10/afteraccess-ict-access-and-use-in-nepal-and-the-global-south/

LIRNEasia. (2021). Access to services during COVID-19 in "Digital India." Available at: https://lirneasia.net/wp-content/uploads/2021/11/COVID-IN_dissemination-deck-full-set-v8.3.pdf

LIRNEasia. (2021c). Digital Sri Lanka during COVID-19 lockdowns. Available at: https://lirneasia.net/wp-content/uploads/2021/12/COVID-LK_dissemination_v7.8.pdf

Little, R. J., & Rubin, D.B., (2002) Statistical analysis with missing data, 2nd edn. John Wiley & Sons, New York

Lu, Y., Ingram, S. & Gillet, D. (2013). A recommender system for job seeking and recruiting website. WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web. 963-966 https://doi.org/10.1145/2487788.2488092

Marinescu, I. (2017). "The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board." Journal of Public Economics 150 (June): 14–29. https://doi.org/10.1016/j.jpube-co.2017.02.012

Matsuda, N., Ahmed, T., & Nomura, S. (2019). Labor market analysis using big data: The case of a Pakistani online job portal. World Bank Policy Research Working Paper, (9063) http://hdl.handle.net/10986/32672

Mckinsey & Company. (2017). India's Labour Market A New Emphasis On Gainful Employment. Available at : https://www.mckinsey.com/~/media/mckinsey/featured%20insights/employment%20and%20growth/a%20new%20emphasis%20on%20gainful%20employment%20in%20india/indias-labour-market-a-new-emphasis-on-gainful-employment.ashx

Messum, D., Wilkes, L., & Jackson, D. (2011). Employability skills: essential requirements in health manager vacancy advertisements. Asia Pacific Journal of Health Management, 6(2), 22-28. Available at https://search.informit.org/doi/pdf/10.3316/ielapa.405517595697217?download=true

Messum, D., Wilkes, L., Peters, K., & Jackson, D. (2016). Content analysis of vacancy advertisements for employability skills: Challenges and opportunities for informing curriculum development. Journal of Teaching and Learning for Graduate Employability, 7(1), 72-86 https://doi.org/10.21153/jtlge2016vol7no1art582

Mezzanzanica, M., & Mercorio, F. (2019). "Big Data for Labour Market Intelligence: An Introductory Guide." European Training Foundation. https://www.etf.europa.eu/en/publications-and-resources/publications/big-data-labour-market-intelligence-introductory-guide

Mhamdi, D., Moulouki, R., El Ghoumari, M. Y., Azzouazi, M., & Moussaid, L. (2020). Job recommendation based on job profile clustering and job seeker behavior. Procedia Computer Science, 175, 695-699.

Ministry of Finance, Nepal. (2021). Economic survey 2020/2021. Available at: https://www.mof.gov.np/uploads/document/file/1633341980_Economic%20Survey%20(Engslish)%202020-21.pdf

Ministry of Labor and Human Resources, Bhutan. (2020). Labour Market Information Bulletin 2020. Available at: https://www.molhr.gov.bt/molhr/wp-content/uploads/2020/11/LMIB-2020.pdf

Moazzem, K. G., Taznur, T. (2021). Impact of the COVID-19 on the Labour Market : Policy Proposal for Trade Union on Employment, Gender and Social Security for Sustainable Recovery. Dhaka: Centre for Policy Dialogue (CPD) and Bangladesh Institute of Labour Studies (BILS). https://cpd.org.bd/wp-content/uploads/2021/08/Impact-of-COVID-19-on-the-Labour-Market-Policy-Proposals-for-Trade-Unions.pdf

Mohan, M.M., Upadhyaya, P. & Pillai, K.R. Intention and barriers to use MOOCs: An investigation among the post graduate students in India. Educ Inf Technol 25, 5017–5031 (2020). https://doi.org/10.1007/s10639-020-10215-2

Monash Data Fluency. (2022). Legal and Ethical Considerations - Python Web Scraping. GitHub. https://monashdatafluency.github.io/python-web-scraping/section-5-legal-and-ethical-considerations/

Moroz, H.E., Naddeo, J.J., Ariyapruchya, K., Jain, H., Glinskaya, E.E., Lamanna, F., Laowong, P., Nair, A., Palacios, R.J., Tansanguanwong, P., Viriyataveekul, S., Walker, T., & Yang, J. (2021). Aging and the Labor Market in Thailand : Labor Markets and Social Policy in a Rapidly Transforming and Aging Thailand (English). Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/428491622713258312/Aging-and-the-Labor-Market-in-Thailand-Labor-Markets-and-Social-Policy-in-a-Rapidly-Transforming-and-Aging-Thailand

Muhleisen, M.B., Zwisler, D.L., & Chacon, R. (2021) Colorado Revises Guidance on Job Posting Requirements. SHRM. https://www.shrm.org/resourcesandtools/legal-and-compliance/state-and-local-updates/pages/colorado-revises-guidance-on-job-posting-requirements.aspx

Mytna Kurekova, L., & Zilincikova, Z. (2018). What is the value of foreign work experience for young return migrants? International Journal of Manpower. http://dx.doi.org/10.1108/IJM-04-2016-0091

National Skills Commission, Australia. (n.d.). Jobs and Education Data Infrastructure (JEDI). Available at: https://www.nationalskillscommission.gov.au/topics/jedi

National Statistics Bureau, Bhutan. (2020). 2020 Labour Force Survey Report Bhutan. Available at: https://www.nsb.gov.bt/wp-content/uploads/dlm_uploads/2021/04/2020-Labour-Force-Survey-Report.pdf

Nitschke, J., O'Kane, L., Taska, Bledi., Hodge & Nyerere. (2021). Big Data for the Labor Market: Sources, Uses and Opportunities. Issues Paper No. 13 APEC Policy Support Unit. Available at https://www.apec.org/publications/2021/12/big-data-for-the-labor-market-sources-uses-and-opportunities

OECD/ADB. (2020). Employment and Skills Strategies in Indonesia, OECD Reviews on Local Job Creation, OECD Publishing, Paris. https://doi.org/10.1787/dc9f0c7c-en.

OECD. (2021). "An assessment of the impact of COVID-19 on job and skills demand using online job vacancy data", OECD Policy Responses to Coronavirus (COVID-19), OECD Publishing, Paris, https://doi.org/10.1787/20fff09e-en

Prasain, S. (2021, June, 17). Tourism is Nepal's fourth largest industry by employment, analytical study shows. Tha Katmandu Post. https://kathmandupost.com/money/2021/06/17/tourism-is-nepal-s-fourth-largest-industry-by-employment-study#:~:text=Money-,Tourism%20is%20Nepal's%20fourth%20largest%20industry%20by%20employment%2C%20analytical%20study,in%20all%20industries%20in%20Nepal

Reserve Bank of Fiji. (2022). Quarterly review: March 2022. Available at: https://www.rbf.gov.fj/quarterly-review-march-2022/

Rivas, R. (2022). More Filipinos join labor force in February 2022 as economy reopens. https://www.rappler.com/business/unemployment-rate-philippines-february-2022/

Swarna, N.R., Anjum, I., Hamid, N.N., Rabbi, G.A., Islam, T., Evana, E.T., et al. (2022). Understanding the impact of COVID-19 on the informal sector workers in Bangladesh. PLoS ONE 17(3): e0266014. https://doi.org/10.1371/journal.pone.0266014

Tas, E. O, Ahmed, T., Matsuda, N; and Nomura, S. (2021). Impacts of COVID-19 on Labor Markets and Household Well-Being in Pakistan: Evidence From an Online Job Platform (English). South Asia Gender Innovation Lab Policy Brief Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/366361617082088695/Impacts-of-COVID-19-on-Labor-Markets-and-Household-Well-Being-in-Pakistan-Evidence-From-an-Online-Job-Platform

Ternikov, A. (2022). Soft and hard skills identification: insights from IT job advertisements in the CIS region. PeerJ Computer Science, https://doi.org/10.7717/peerj-cs.946

Tokona te Raki. (2020). Regional & National Employment & Skills Report. Available at: http://www.maorifutures.co.nz/wp-content/uploads/2020/08/Te-Kete-Pukenga-Wh%C4%81nau-Report.pdf

Turrell, A., Speigner, B., Djumalieva, J., Copple, D., and Thurgood, J. (2019). “Transforming Naturally Occurring Text Data Into Economic Statistics: The Case of Online Job Vacancy Postings.” Working Paper 25837. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w25837.

Turrell, A., Speigner, B., Copple, D., Djumalieva, J., and Thurgood, J. (2021). “Is the UK’s Productivity Puzzle Mostly Driven by Occupational Mismatch? An Analysis Using Big Data on Job Vacancies.” Labour Economics 71 (August): 102013. https://doi.org/10.1016/j.labeco.2021.102013

UN Women. (2022). Gender disparities and labour market challenges: The demand for women workers in Sri Lanka. Available at: https://asiapacific.unwomen.org/sites/default/files/2022-03/lk-Gender-Disparities-and-Labour-Market-Challenges_Full-Report.pdf

UNCTAD. (2017). Bhutan Rapid eTrade Readiness Assessment. Available at: https://unctad.org/system/files/official-document/dtlstict2017d1_en.pdf

UNDP. (2021). Digital Jobs in Bhutan: Future Skilling and Demand Creation. Available at: https://www.undp.org/bhutan/publications/digital-jobs-bhutan-future-skilling-and-demand-creation

UNICEF. (2019). Developing skills in youth to succeed in the evolving South Asian economy. Bhutan country report. Available at: https://www.unicef.org/rosa/media/4486/file/Bhutan%20CR.pdf

Vacancy Survey. (n.d.). Office for National Statistics. https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/vacancysurvey

Van Loo, J., Pouliakas, k. (2020). “Cedefop and the Analysis of European Online Job Vacancies.” In The Feasibility of Using Big Data in Anticipating and Matching Skills Needs. Geneva, Switzerland: ILO. Available at https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_ent/documents/publication/wcms_759330.pdf

Verick, S. (2014). Female labor force participation in developing countries. IZA World of Labor. Available at: https://wol.iza.org/articles/female-labor-force-participation-in-developing-countries/long

Wignaraja, G. (2018). Escaping the paradox of slow growth and labour scarcity in Sri Lanka. https://blogs.lse.ac.uk/southasia/2018/08/15/escaping-the-paradox-of-slow-growth-and-labour-scarcity-in-sri-lanka/

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy. Work, employment & society : a journal of the British Sociological Association, 33(1), 56–75. https://doi.org/10.1177/0950017018785616

World Bank Group; Ministry of Labor and Human Resources, Royal Government of Bhutan. (2016). Bhutan's Labor Market : Toward Gainful Quality Employment for All. Washington, DC: World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/25703 License: CC BY 3.0 IGO.

World Bank. (2018). Sri Lanka Development Update, June 2018 : More and Better Jobs for an Upper Middle-Income Country. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/29927 License: CC BY 3.0 IGO.

World Bank. (2018a) Indonesia's Critical Occupations List 2018. Available at: https://documents1.worldbank.org/curated/en/763611585857010121/pdf/Indonesias-Critical-Occupations-List-2018-Technical-Report.pdf

World Bank. (2019). Malaysia's 'Critical Occupations List' is an Innovative Tool for Preparing Workers for the Jobs of the Future: World Bank. Available at: https://www.worldbank.org/en/news/press-release/2019/09/12/malaysias-critical-occupations-list-is-an-innovative-tool-for-preparing-workers-for-the-jobs-of-the-future-world-bank

World Bank Group. (2019). Systematic Country Diagnostic of the Philippines : Realizing the Filipino Dream for 2040. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/32646 License: CC BY 3.0 IGO.

World Bank. (2020). Philippines Digital Economy Report 2020 : A Better Normal Under COVID-19 - Digitalizing the Philippine Economy Now. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/34606 License: CC BY 3.0 IGO.

World Bank. (2021). The World Bank In India. https://www.worldbank.org/en/country/india/overview#1

World Bank. (2022). World Development Indicators [various]. Retrieved from https://databank.worldbank.org/source/world-development-indicators

World Bank. (2022a). COVID-19 in South Asia : An Unequal Shock, An Uncertain Recovery - Findings on Labor Market Impacts from Round 1 of the SAR COVID Phone Monitoring Surveys. Washington, DC. Available at: https://openknowledge.worldbank.org/handle/10986/37320

Xu, T., Zhu, H., Zhu, C., Li, P., & Xiong, H. (2017). Measuring the Popularity of Job Skills in Recruitment Market: A Multi-Criteria Approach. https://www.researchgate.net/publication/321719277_Measuring_the_Popularity_of_Job_Skills_in_Recruitment_Market_A_Multi-Criteria_Approach

Yamauchi, F., Nomura, S., Imaizumi, S., Areias, A. C., & Chowdhury, A. R. (2018). Asymmetric information on noncognitive skills in the Indian labor market: an experiment in online job portal. World Bank Policy Research Working Paper, (8378) http://hdl.handle.net/10986/29558

Zainudeen, A. (2021). Impacts of the COVID-19 pandemic on income and earning opportunities: Findings from nationally representative surveys in India and Sri Lanka. Available at: https://connected2work.org/blog/impacts-of-the-COVID-19-pandemic-on-income-and-earning-opportunities-findings-from-nationally-representative-surveys-in-india-and-sri-lanka/

Zhang, Y., Yang, C., & Niu, Z. (2014). A research of job recommendation system based on collaborative filtering. In 2014 Seventh International Symposium on Computational Intelligence and Design (Vol. 1, pp. 533-538). IEEE.