

# Toward Labor Market Policy 2.0

## The Potential for Using Online Job-Portal Big Data to Inform Labor Market Policies in India

*Shinsaku Nomura*

*Saori Imaizumi*

*Ana Carolina Areias*

*Futoshi Yamauchi*



**WORLD BANK GROUP**

Education Global Practice Group

February 2017

## Abstract

Economists and other social scientists are increasingly using big data analytics to address longstanding economic questions and complement existing information sources. Big data produced by online platforms can yield a wealth of diverse, highly granular, multidimensional information with a variety of potential applications. This paper examines how online job-portal data can be used as a basis for policy-relevant research in the fields of labor economics and workforce skills development, through an empirical analysis of information generated by Babajob, an online Indian job portal. The analysis highlights five key areas where online job-portal data can contribute to the development of labor market policies and analytical knowledge: (i) labor market monitoring and analysis; (ii) assessing demand for workforce skills; (iii) observing job-search behavior and

improving skills matching; (iv) predictive analysis of skills demand; and (v) experimental studies. The unique nature of the data produced by online job-search portals allows for the application of diverse analytical methodologies, including descriptive data analysis, time-series analysis, text analysis, predictive analysis, and transactional data analysis. This paper is intended to contribute to the academic literature and the development of public policies. It contributes to the literature on labor economics through application of big data analytics to real-world data. The analysis also provides a unique case study on labor market data analytics in a developing-country context in South Asia. Finally, the report examines the potential for using big data to improve the design and implementation of labor market policies and promote demand-driven skills development.

---

This paper is a product of the Education Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [snomura@worldbank.org](mailto:snomura@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# **Toward Labor Market Policy 2.0: The Potential for Using Online Job-Portal Big Data to Inform Labor Market Policies in India<sup>1</sup>**

**Shinsaku Nomura\*, Saori Imaizumi, Ana Carolina Areias, Futoshi Yamauchi**

The World Bank

JEL Classification: C55, C81, J23, J24, J31, J71

Key words: labor market analytics, big data, wage, gender, skills demand, behavioral, forecasting

---

<sup>1</sup> This work was supported by the World Bank Strategic Research Program Trust Fund and the Jobs Umbrella Trust Fund, which is supported by the Department for International Development/UK AID, and the Governments of Norway, Germany, Austria, the Austrian Development Agency, and Swedish Development Agency SIDA. The authors are grateful to Babajob, with special thanks to Sean Blagsvedt and Vir Kashyap for their continuing collaborations and generous data sharing, and to John Gibbons for operational task management. John D. Blomquist has provided continuous guidance to the team since the early stage of this collaborative research. The paper has been prepared to contribute to the Skill India Mission Operation project and skills development sector knowledge in India, and the team is grateful to Shabnam Sinha and Muna Meky for their advice. The team benefited from technical discussions with Omar Arias, Mattias Lundberg and the World Bank India education team at various stages of this research, and overall guidance by Keiko Miwa.

\* Corresponding author: The World Bank, Education Global Practice. Washington, D.C.: [snomura@worldbank.org](mailto:snomura@worldbank.org).

## **1. Introduction**

Education and skills development are known to contribute to increased labor productivity, poverty reduction and economic growth. As a result, many countries around the world prioritize education as a core pillar of human capital development and economic growth. According to the OCED, “skills have become the global currency of the 21st century” (OECD, 2012). Recent studies suggest that the quality of education and skills training is more critical to economic development than the quantity (Hanushek and Woessmann 2008, 2011). This new evidence, combined with substantial progress in expanding and improving primary education, has helped to shift the focus of human capital formation in the international development discourse from achieving universal primary education under the United Nations Millennium Development Goals to promoting skills development and job creation under the Sustainable Development Goals (SDGs). The SDGs highlight the importance of improving access to high-quality, employer-relevant skills training and advancing opportunities for employment and entrepreneurship. The World Bank has established a human capital formation framework that regards skills development as a lifelong process (World Bank, 2010).

Technological advances and improved internet connectivity have spurred the creation of jobs requiring new and diverse skills. In an increasingly digital global economy, many functions once thought to be the exclusive domain of humans, such as driving or writing, are increasingly being carried out by computers. Going forward, policy makers will need to focus on developing skills that are less vulnerable to automation (World Bank 2016). These include non-routine and higher-order skills such as complex information processing, critical thinking, problem solving and reasoning. Studies have shown that workers who have developed such skills earn 25 to 40 percent more than their peers with similar education levels when performing traditional tasks (World Bank 2016).

In addition to the overarching challenge of human capital development, mismatches between workforce skills and employer demand remain a serious obstacle to economic growth. Skills mismatches can occur at the macro-level, when the overall quantity and/or quality of workforce skills does not correspond with the demands of the labor market, or at the micro-level, when individual employers cannot identify appropriately skilled hires. Macro-level skills mismatches often occur when a country’s labor market and education and training systems are only weakly connected. Despite continuous efforts to move toward a more demand-driven skills-development framework, international examples of successful demand-driven systems are limited, as countries often lack detailed real-time data on employment dynamics. Information and coordination failures in the labor market can also make it difficult for employers to identify qualified candidates. While some successful training institutions effectively involve local business partners, few have

access to systematic assessments of labor demand or the supply of workforce skills. Moreover, despite the large size and economic importance of the informal sector in many developing countries and emerging markets, assessments of supply and demand in the informal labor market are often minimal or wholly absent. As a result, strategies to improve skills matching and keep training providers apprised of evolving demand conditions in the labor market remain important challenges.

There is emerging evidence that analyses of online job-search engines and other employment services can be used to address information asymmetry and coordination failures in the labor market. Global online job portals such as LinkedIn, Indeed, Monster and Career Builder are bridging the gap between employers and job seekers in many countries. In areas where the use of such websites is still limited, local online job portals often provide vacancy announcements for both the formal and informal sectors and help expand access to job information for those with internet connectivity. These online platforms, whether global or local, provide a rich and continuous stream of labor market data that has remained largely untapped by policy makers.<sup>2</sup>

In the field of labor economics, the prevalence of big data from online job portals is increasing, and various researchers have explored different methodological approaches to analyzing this data. For instance, Kureková et al. (2015) conducted a literature review in an effort to identify the advantages and disadvantages of using online job data and concluded that it represented a rich source of information that should be exploited despite issues with the representativeness of online job-portal data. Labor economics has traditionally relied on sample survey data, such as workforce and household surveys. Such surveys are often infrequently conducted, and their scope and coverage can entail important limitations. By contrast, big data has the potential to generate multidimensional, granular, real-time information, which can be used to provide new insights into longstanding economic questions, enhance economic research and aid public policy development.

## **2. Objectives**

This paper discusses potential applications of online job-portal data for academic research and policy formulation through an empirical study of data produced by Babajob, an online job-matching platform in India. The paper is designed to both contribute to the academic literature and to support the development

---

<sup>2</sup> For example, LinkedIn, a social networking service for businesses and professionals launched in 2003, had data on 433 million individuals as of Q1 2016. The site is available in over 200 countries worldwide and in 20 different languages. The data obtained through LinkedIn have been primarily used for business purposes, rather than for policy formulation, though various industry- and skills-focused analyses have been published on LinkedIn's official blog. Tambe (2014) used skills data from LinkedIn to measure employers' investment in big-data-related human resources management.

of more effective labor market policies in India and elsewhere. Much of the current literature on big data analytics in economic research is limited to methodological reviews and analyses of data sources, and does not examine real-world applications. This report not only reviews the prospective research applications of big data analytics, but also demonstrates their utility through an analysis of actual Babajob data. By using data from India, the study provides a unique case study in a developing country context. Unlike data from online job portals in the United States or Europe, labor market data in India is not regularly updated, making it difficult to assess changes in the labor market. This paper can help shed new light on labor market dynamics in India and expand the possibilities for using big data analytics to enhance labor market policies and promote demand-driven skills development.

The paper discusses five potential applications of big data in formulating and refining labor market policies: (i) conducting labor market monitoring and analysis; (ii) assessing demand for workforce skills; (iii) observing job-search behavior and improving skills matching; (iv) predictive analysis of skills demand; and (v) experimental studies. The paper applies the following analytical methodologies: descriptive-data analysis, time-series analysis with spatial information, text analysis, predictive analysis and transactional-data analysis.

### **3. Literature on Economic Research Using Big Data**

The term “big data” first appeared in the early 2000s, as a surge in data processing capacity, internet use and detailed data collection created an unprecedented wealth of granular, real-time information on the behavior of individuals and organizations. “Volume, variety and velocity” are the three defining properties of big data (Laney 2001). Big data analytics was initially leveraged by firms seeking to adapt their business models and day-to-day decision making to reflect emerging data on consumer behavior. Since then, big data analytics has been increasingly utilized as a critical management tool and knowledge source (McAfee and Brynjolfsson, 2012), occupying a space between academic and applied knowledge (Savage and Burrows 2007, 2009; Taylor, Schroeder, and Meyer 2014).

More recently, the use of big data has expanded rapidly in economics and other social sciences. The advent of big data has already prompted novel research designs across a range of topics and refined the measurement of economic effects and outcomes. Over time, these data are likely to affect the types of questions economists pose, allowing for greater attention to variation across populations and facilitating analyses across a broader range of economic activities and interactions. Einav and Levin (2013) point to three unique characteristics of big data that can help contribute to the economic literature: (i) the real-time

nature of the data, (ii) the scale, multidimensionality and granularity of the information collected; and (iii) its ability to capture actual human behavior recorded in real-world transactions.

The emergence of big data has also enabled researchers to examine aspects of the labor market that have historically been difficult to analyze (Horton and Tambe, 2015). For instance, using data from the predominately US-focused online job portal “Career Builder,” Marinescu and Rathelot (2015) assessed the effects of geography on worker application patterns. Their research concluded that while job seekers typically do not want to apply for jobs far from where they live, the effects of geography are too small to explain much of the frictional unemployment in the labor market. This type of analysis would have been impossible without big data analytics. Other studies using data from job portals have been conducted in the US, Europe and middle-income countries like China and the Slovak Republic. Moreover, social networks such as LinkedIn, Facebook and Twitter, employment-information websites such as Glassdoor, and trend-information providers like Google Trends, are also potential all new data sources for labor market analysis (Lenaerts, Bevlavy and Fabo, 2016).

The establishment and rapid expansion of online job platforms have dramatically reduced the cost of information access for both job seekers and employers. Kroft and Pope (2014) and Mang (2012) argue that there has been a large-scale shift from posting job listings in newspapers and other print media to websites and that this process has lowered the cost of acquiring employment information. Moreover, several studies from developed countries have found that the use of employment websites has reduced unemployment rates (Beard et al., 2012; Kuhn and Mansour, 2014), though other studies do not find strong evidence of an impact on wages (Kuhn and Mansour, 2014; Shahiri and Osman, 2014). However, as Shahiri and Osman (2014) point out, many people in developing countries—including both employers and job seekers—cannot take advantage of employment websites due to limited internet infrastructure, high user fees, and/or a lack of knowledge regarding information technology. Studies have also found that greater cellphone use improves employment outcomes in developing countries (Klonner and Nolen, 2010 for South Africa; Aker, 2011 for Niger; Burga and Barreto, 2014 for Peru), lending further support to the conclusion that information constraints are a significant, though often unrecognized, source of inefficiency in the labor market.

There are important challenges involved in using big data for labor market analysis. These include accessing and analyzing data in ways that respect the privacy and confidentiality of individuals, designing creative and scalable approaches to summarizing, describing and analyzing unstructured data sets, and dealing with the partial representativeness of the data. Methodological challenges and the ownership of data sets are also important considerations. One of the principal critiques of big data analytics is that the underlying empirical

micro-processes that lead to the emergence of the typical network characteristics of big data are not well understood (Snijders, Matzat, and Reips 2012).

#### **4. The Indian Labor Market, Skills Development and Babajob**

India has made impressive progress in economic growth and poverty reduction over the past few decades. India's gross domestic product (GDP) grew at an average rate of 7.3 percent per year between 2007 and 2012. This contributed to a substantial decline in the incidence of poverty, and an estimated 138 million people rose above the poverty line during the period. Despite its robust growth, India still faces major challenges in improving worker productivity and ensuring that the supply of workforce skills is consistent with labor market demand.

The Indian labor force is large, mostly informal and relatively young. India's population reached 1.295 billion in 2014, including 497 million workers.<sup>3</sup> The size of the labor force has been expanding at an annual net growth rate of 4.2 million for the past 10 years. The labor force participation rate is 54 percent, with a relatively high participation rate among men (80 percent) and low rate among women (26 percent). Only 16 percent of the labor force is engaged in wage employment (18 percent of male workers and 12 percent of female workers), and a large majority work in the informal sector. Moreover, 54 percent of the country's population is under 25 years of age. While the country's relatively young population has the potential to yield significant demographic dividends, ensuring sufficient employment opportunities for a young workforce also presents a critical challenge for the government. India faces a lack of highly trained workers and a large share of unskilled youth. As a result, job creation and skills development are critical priorities. To address these challenges, the government launched the National Policy for Skill Development and Entrepreneurship in 2015, and its 12<sup>th</sup> Five-Year Plan set a goal of training 400 million workers by 2022.

India's national labor market data are mainly collected through National Sample Surveys, some rounds of which have dedicated sets of employment questions. The Central Statistical Organization also conducts an Annual Survey of Industries and quarterly employment surveys designed to track industrial and workforce development trends in selected economic sectors. External surveys, such as the International Finance Corporation's Enterprise Surveys, periodically analyze India's business environment, and especially its formal sector (IFC 2014). Sample surveys of households and industries are typically conducted by the government and have traditionally served as the primary data source for understanding labor market trends and skills demand in India.

---

<sup>3</sup> World Bank World Development Indicators website.

Online job portals emerged in India in the late 1990s, but only began to flourish in the past decade as mobile phone and internet use became more widespread and social networking platforms expanded. The share of mobile phone subscribers more than tripled from 20 percent of the population in 2007 to 70 percent in 2014, and the number of fixed broadband subscriptions increased eightfold during the same period. There are now about 20 job search portals, many of them focusing solely on the Indian labor market.<sup>4</sup> Babajob, established in 2007, is one of the country's leading job-matching websites. Between 2007 and 2015 more than 858,000 jobs were posted on the site, and over 240,000 employers and 4.5 million job seekers were registered. Babajob is unique in matching workers to potential employers in both the formal and informal sectors. In order to better reach disadvantaged populations, Babajob provides a variety of access options including standard websites, mobile sites, interactive voice response (IVR), text messaging and web applications.<sup>5</sup>

A substantial share of the jobs listed on Babajob are entry level, with the largest number of listings in 2015 classified as clerical support.<sup>6</sup> 70 percent of the jobs advertised were based in the country's 10 most populous cities: Bangalore, Delhi, Mumbai, Chennai, Hyderabad, Pune, Kolkata, Thane, Patna, and Lucknow. The average salary offers for all jobs on the website was 13,182 rupees (Rs.) per month,<sup>7</sup> with the average for professional-level jobs (Rs. 14,900) being 17 percent higher than the average for non-professional jobs (Rs. 12,739). By city, average listed salaries for professional jobs ranged from Rs. 16,970 in Mumbai to Rs. 12,757 in Patna (a 33 percent difference), while the average for non-professional jobs ranged from Rs. 14,184 in Delhi to Rs. 10,742 in Patna (a 32 percent difference).

---

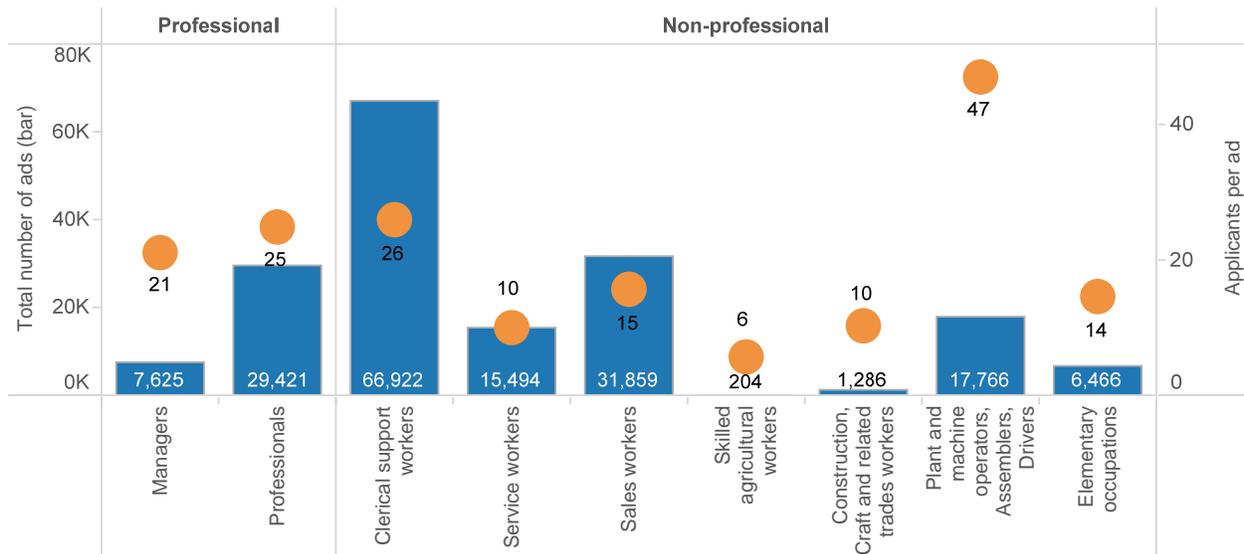
<sup>4</sup> Though not exhaustive, 22 firms were identified by compiling a list from government agencies and search engines. 77 percent of identified firms provide job-search services, while a few provide job-matching services. Some platforms focus on entry-level jobs, while others primarily advertise technology or senior management-level jobs. Job-matching platforms use different techniques to match workers with job opportunities, including leveraging social networks, providing curated job information based on an individual's profile and connecting local recruiters with candidates. 68 percent of the researched platforms focused on jobs located in India, with some portals concentrating on specific cities.

<sup>5</sup> The Babajob platform works to connect job seekers with job opportunities and employers with candidates for their positions. Job seekers can create a profile on Babajob, search and apply to jobs online or offline, all for free. Employers can create profiles on Babajob and post their hiring requirements for free in a service that resembles online job classifieds, or they can opt for the paid, premium service (*RapidHire*) that offers a facilitated hiring experience. All job posts are live for 90 days, with *RapidHire* jobs promoted more heavily on the site for a period of time depending on the type of plan opted for (e.g. the basic offering promotes jobs for 15 days). *RapidHire* also includes other services, such as additional screening, a recruitment support executive, the ability to 'unlock' more information about candidates, and SMS promotion of the available job to relevant job seekers within a certain radius.

<sup>6</sup> These definitions are based on the International Standard Classification of Occupations (ISCO). Professional-level positions include managers and technicians, professionals and associate professionals. Non-professional occupations include clerical workers, service workers, sales workers, skilled agricultural workers, construction and craft workers, plant workers, drivers and elementary occupations.

<sup>7</sup> Using an exchange rate of US\$1=Rs. 67 (the average rate during the first half of 2016) this amount is equivalent to about US\$197.

**Figure 1: Number of Job Listings on Babajob and Average Number of Applications per Listing by Occupational Category, 2015**



Source: Authors' calculations using Babajob data

## 5. Using Big Data to Analyze Labor Market Dynamics in India: the Case of Babajob

A review of the existing literature and a thorough analysis of Babajob data revealed five ways in which online job-portal data can be used to formulate and refine labor market policies. These include (i) labor market monitoring and analysis, (ii) assessing demand for workforce skills, (iii) observing job-search behavior and improving skills matching, (iv) predictive analysis of skills demand, and (v) experimental studies.

### 5.1. Labor Market Monitoring and Analysis

Online job-portal data can be used to monitor and analyze labor market trends in real time. Big data can complement official government statistics and other traditional forms of employment information-gathering, such as sample-based labor force and enterprise surveys, to provide richer and more granular data. Labor force surveys capture information on the personal characteristics, wages and experience levels of employees, as well as employer characteristics such as firm size and workforce composition. Enterprise surveys focus on formal-sector business and production activities and assess employee productivity. Both labor force and enterprise surveys tend to be conducted infrequently and focus on a small number of variables. A key advantage of online job-portal data is that it allows for a much larger sample to be monitored in real time, enabling analysts and policy makers to develop a more thorough and timely understanding of evolving conditions in the labor market.

In addition to the potential for real-time monitoring, the granularity and uniqueness of data captured by online job portals offer unique advantages over traditional survey instruments. Online job advertisements can provide highly granular data on labor market characteristics and demand dynamics at a specific time or location, or for a specific type of occupation. For example, while official statistics collect data on the current level and distribution of wages, online job-portal data focus solely on wage offers, which can help identify emerging trends in the demand for specific skills, the seasonality of hiring practices and firms' production expectations for specific jobs. The data allow for both a dynamic and static evaluation of the labor market, as well as an examination of specific market segments. Information on the historical growth and spatial distribution of job openings, wage trends and patterns, and competition among job seekers can help policy makers understand how the labor market is evolving in different places and over time.

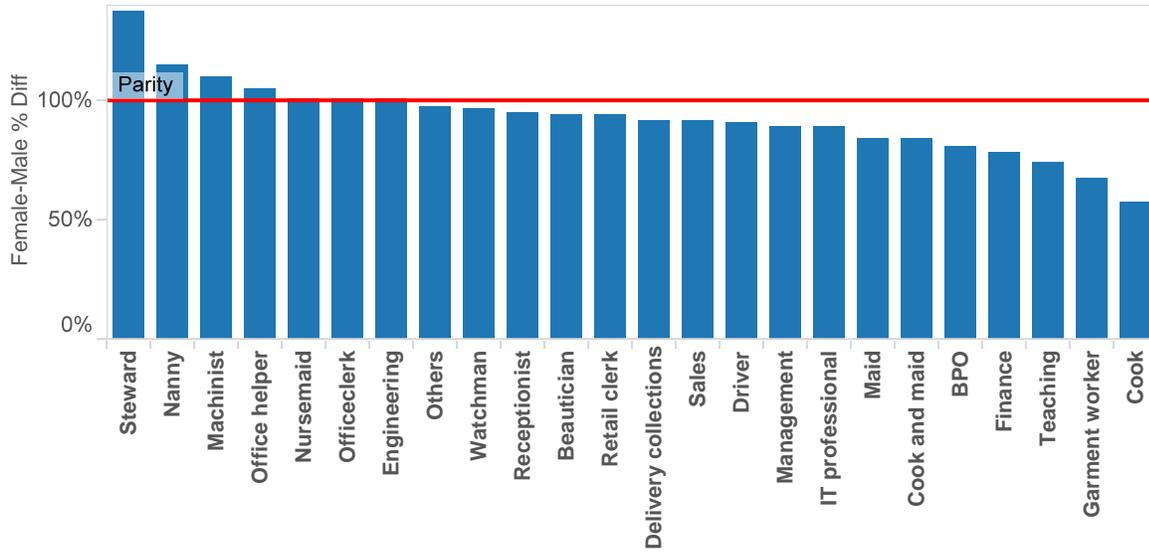
However, there are important caveats regarding the representativeness of online job-portal data. In order to ensure that findings are interpreted accurately, it is crucial to specify exactly what the data are representing and to clearly define their applicability to the broader labor market. Limitations on representativeness can be addressed by statistically weighting the data according to the industry structure derived from labor force surveys, focusing on segments of the labor market in which coverage bias is less of a problem, and using diverse data sources to conduct parallel analyses and confirm the robustness of findings (Kureková, Beblavý, and Thum-Thysen, 2015).

### ***Application of the Analysis to Babajob Data***

As noted above, online job-portal data can facilitate an in-depth analysis of specific social groups and segments of the labor market. In this study, Babajob data was used to analyze gender differences in wage offers for similar positions. Women constitute 26 percent of the Indian workforce, and gender disparities in workforce participation rates are an important development issue (United Nations, 2012). An analysis of wage offers for 858,000 jobs posted to Babajob by 240,000 unique employers between 2007 and 2015 reveals how these offers evolved over time and varied by location, occupation and contract type. An econometric analysis found that the average wage offers for positions that specified a male hire were 7.1 percent higher than those that were open to both genders, while wage offers for positions that specified a female hire were 16.2 percent lower. This analysis controlled for a number of other factors that might affect wage rates, including employment period, location, occupation and contract type. Among occupations that were frequently offered to both male and female workers, cooks, garment, and teaching jobs exhibited relatively high wage differentials, while stewards, nanny, machinist and office helpers exhibited relatively small differences (Figure 2). Among the 20 cities with the largest number of job listings, Ahmedabad,

Ranchi, and Thane evinced the widest gender wage gaps, while wage gaps in Gurgaon, Noida, and Indore were narrower (Figure 3).

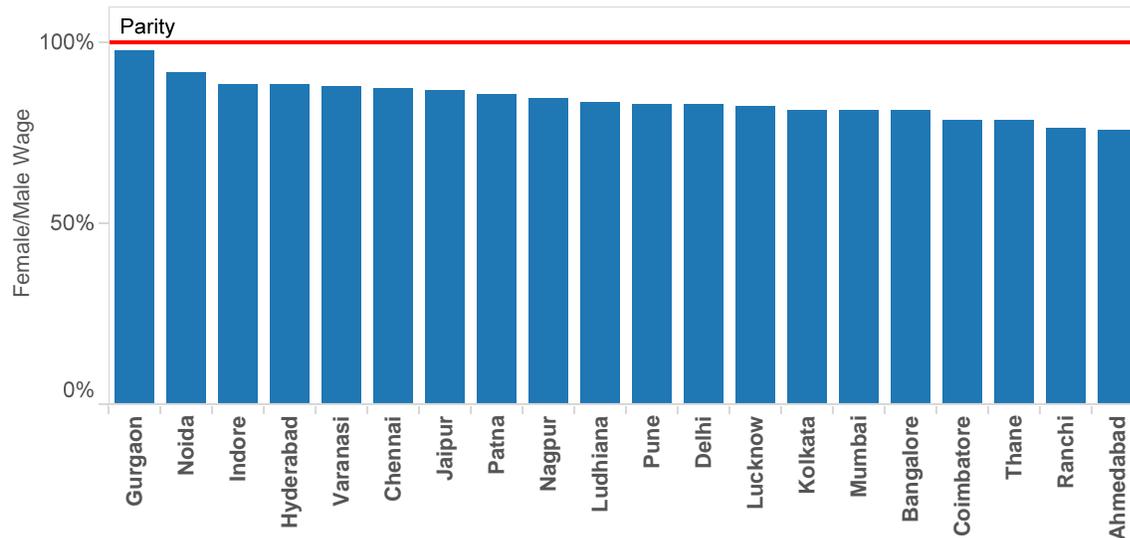
**Figure 2: Gender Wage Gap by Occupation**



Source: Authors' calculations using Babajob data

Note: Average wage offered for female-specified job postings as a percentage of male-specified postings after controlling for other job characteristics. Job category refers to the 26 occupational categories used by Babajob.

**Figure 3: Gender Wage Gap by Location**



Source: Authors' calculations using Babajob data

Note: Average wage offered for female-specified job postings as a percentage of male-specified postings after controlling for other job characteristics.

Gender wage gaps are common in many countries.<sup>8</sup> However, given India's low female workforce participation rate, the gender disparity in wage offers is particularly critical. A key finding of the analysis was the variation in gender differentials by location and occupation type. Locations that exhibit lower wage offer differentials such as Gurgaon, Lucknow and Noida may be more conducive to women entering the labor market. Moreover, occupations that are commonly held by female workers do not necessarily have smaller wage gaps, particularly professional jobs. While more investigations will be needed, more equitable wage structures may be an important labor market condition for encouraging women's work.

## **5.2. Assessing Demand for Workforce Skills**

The contemporary dialogue on workforce development policy tends to focus on demand-driven skills development. In cases where a growing number of educated and trained workers are unemployed or underemployed, the most likely explanation is that the particular skills they provide are not in high demand. While some educational institutions work closely with specific industries to ensure that their graduates possess the skills employers are looking for, overall trends in labor market demand may not be effectively communicated to existing or incoming workers.

<sup>8</sup> See, for example, a 2005 meta-analysis by Weichselbaumer and Winter-Ebmer,

Information on employer demand is most often collected through subjective surveys. Cunningham and Villasenor (2016) reviewed 27 studies from various countries that attempted to identify the skills demanded by employers. They separated skills into four categories—socio-emotional, high-order cognitive, basic cognitive and technical—and found that employers tended to place greater emphasis on socio-emotional skills and higher-order cognitive skills. While employer surveys can help sharpen the focus of education and skills training, the subjective and qualitative nature of employers’ responses can make it difficult to assess the relative importance of different skills and skill combinations (Rutkowski, 2010).

Big data from online job portals can help quantify employer demand through text analysis. Researchers use text analysis to investigate large amounts of textual information and systematically identify its properties. They analyze the frequency of keyword usage, identify communication content and categorize the structure of the text. Using text analysis to examine labor market dynamics offers a number of advantages over more traditional approaches. Online job postings may reflect employers’ desired qualifications and skills more accurately than survey responses. Unlike printed advertisements, employers posting job offers on websites do not have to pay by the word and thus can provide more detailed information on the knowledge and skills they require (Gallivan et al. 2004). These descriptions, in turn, can help highlight the skills demanded by different occupations and in different locations, as well as the wage discrepancies associated with various skill sets. Traditional sample-based labor force surveys cannot capture this type of information.

Beblavý, Fabo and Lenaerts (2016a) examined two million job advertisements published on a US-based online job portal called Burning Glass. They grouped the skills and qualifications required in the postings into several categories, such as education and formal qualifications, cognitive skills, non-cognitive skills and experience, and then broke these categories down into sub-categories. They found that while employers placed a high value on both cognitive and non-cognitive skills, the emphasis on each in job postings varied by employee skill level. While 25 percent of all job advertisements highlighted computer skills, percentages for postings targeting low-, medium- and high-skilled workers varied from 9 percent to 21 percent to 39 percent, respectively. Similarly, while 45 percent of job postings demanded non-cognitive service skills, the share of postings for low-, medium- and high-skilled workers ranged from 21 percent to 47 percent to 45 percent, respectively. Beblavý, Fabo and Lenaerts (2016b) analyzed job ads from the Czech Republic, Hungary, Poland and the Slovak Republic to assess the demand for foreign language skills. Focusing on 59 occupation types, they found that foreign-language skills were in high demand in these countries, with the share of postings that mentioned them ranging from 28 percent in the Czech Republic to 64 percent in Poland. English was the most-requested language and was mentioned in 52 percent of vacancies.

#### ***Application of the Analysis to Babajob Data***

This study used Babajob data to identify the skills demanded by employers in two ways. The first method was to review employer qualification screening questions, and the second was to apply text analysis to capture the demand for skills as expressed in the text of job descriptions. Figure 4 presents the most commonly asked screening questions by occupation level.<sup>9</sup> Four broad categories of screening questions were identified: (i) skills; (ii) personal characteristics; (iii) work conditions and benefits; and (iv) other. The skills category includes overall education level and academic credentials, as well as computer skills, language skills and relevant occupational experience. The work conditions and benefits category includes location or distance to work, salary, and the proposed terms of employment. The “other” category includes uncommon and job-specific questions.

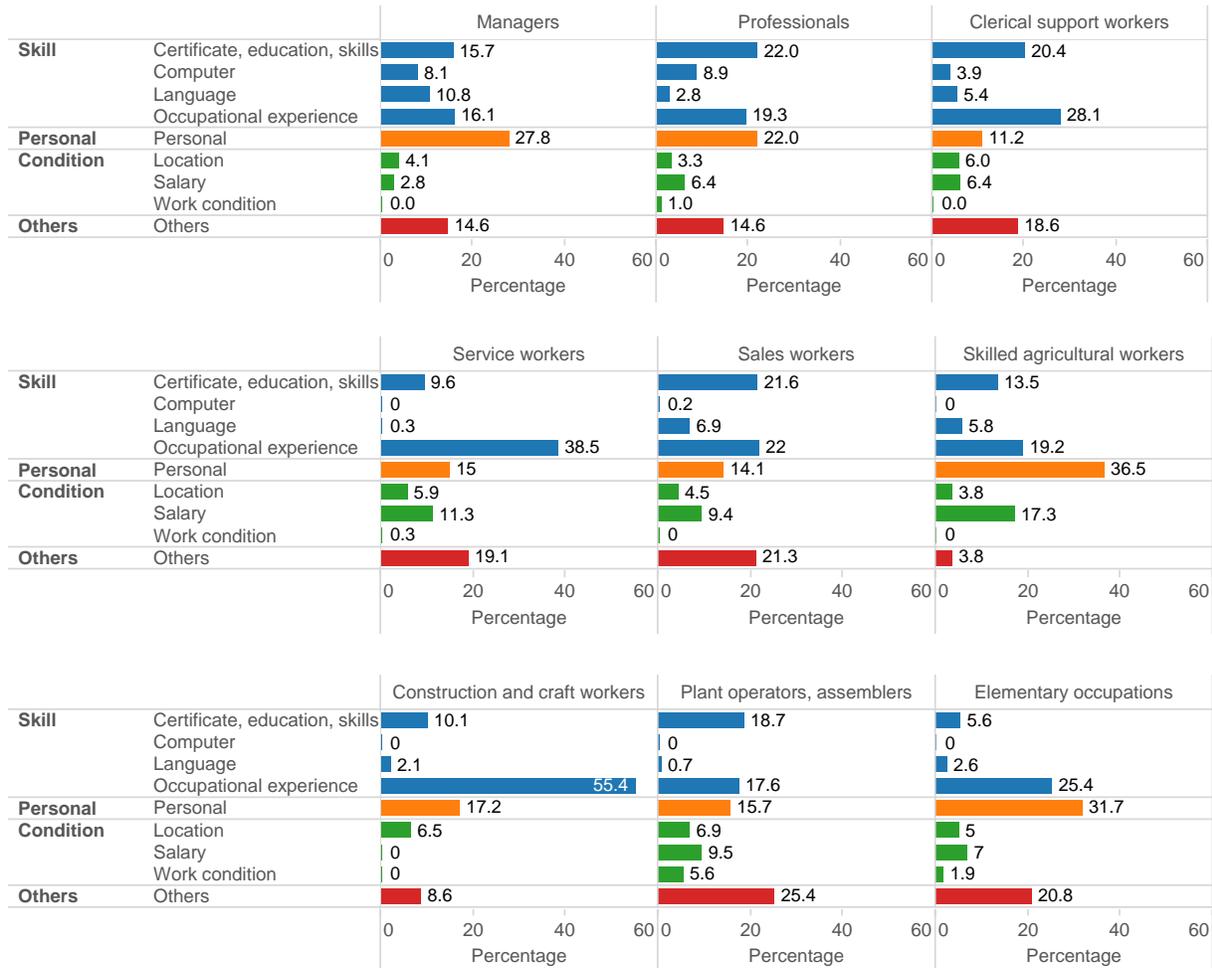
The data reveal that skills-related questions are the most common category of screening filter, though the skills requested differ by occupation level. Skills-related questions appear in about 50 percent of qualifications screenings for most occupation levels, including managerial (50.7 percent) and professional positions (47.9 percent). However, while employers searching for professionals and associate professionals tended to ask about education levels and specific technical skills, occupational experience appears to be more relevant for clerical staff, service and sales workers, and construction and craft workers. Language skills, most often English and Hindi, came up more frequently for managerial and sales positions.

The relatively high frequency of personal information questions demonstrates the importance employers place on collecting basic information, especially for service-related hires in the informal sector, such as maids, cooks and gardeners. Personal information inquiries include questions on the applicant’s gender, age, police verification certificates, as well as requests for resumes and photos. The applicant’s acceptance of the stated employment terms, including the proposed salary and work hours, is another important filter for recruitment decisions.

---

<sup>9</sup> Occupational levels are defined according to the ISCO classification. Employers can either use standard questions or manually enter their own questions.

**Figure 4: Shares of Commonly Asked Screening Questions by Occupational Type**



Source: Authors' analysis using Babajob data (Jan – Oct, 2015).

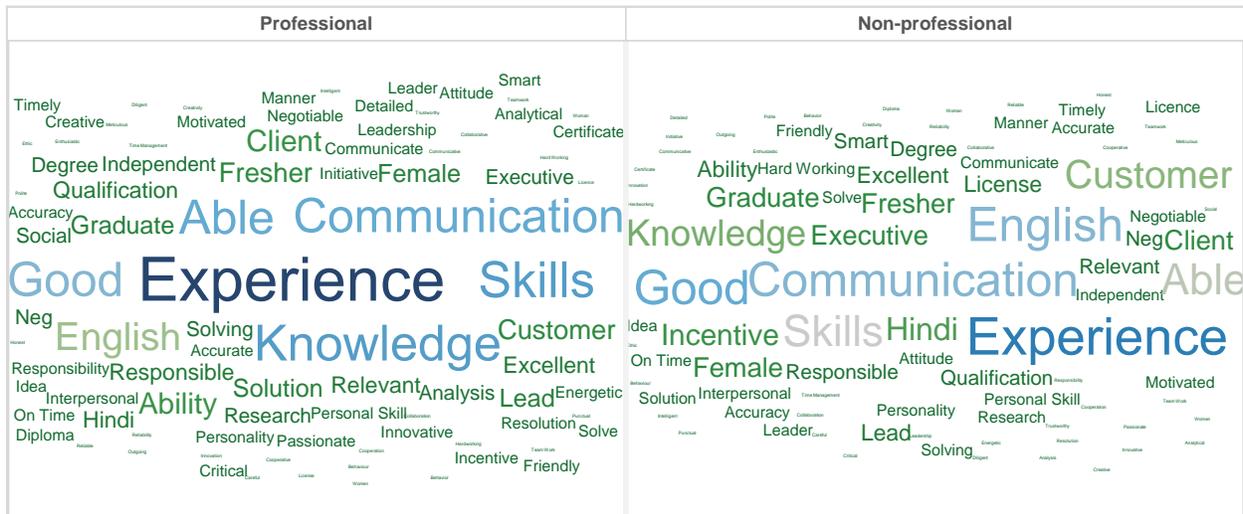
Text analysis can provide additional insight into employer preferences beyond minimum qualification requirements, and it can be especially useful in understanding the cognitive and non-cognitive skills that employers demand for each position. Employers rarely ask job seekers about their non-cognitive skills directly, since there are few objective measures for this skill type. While information on demand for these skills may be obtained from interviews with employers and surveys, an analysis of job advertisements may be more accurate, as it reflects the real-world behavior of employers.

Figure 5 illustrates the skill- and qualification-related words that appeared most frequently in the job descriptions posted to Babajob in 2014. Figure 6 shows the proportion of advertisements using the clusters of these keywords. Overall, skill- and qualification-related words appear more often in professional-level job descriptions. For both professional and non-professional occupations, the words “experience” and

“communication” are most frequently used. The words “English” and “customer/client” are also commonly used, but are more prevalent among non-professional occupations. Keywords related to language skills and customer care appear more frequently in advertisements for non-professional jobs, while keywords related to problem-solving, leadership, analytical skills, work ethic, reliability, creativity and personality attributes tend to appear more often in professional-level job descriptions.

The data suggest that employers have more sophisticated and detailed expectations for professional-level hires in terms of non-cognitive and higher-order cognitive skills. However, the data also show that only a relatively small percentage of job advertisements highlight required non-cognitive skills. This finding differs from that of previous analytical work in India (Blom and Saeki, 2011) and in other countries (Cunningham and Villasenor, 2016), which reported employers’ emphasis on non-cognitive skills. This could be explained that employers may choose not to highlight non-cognitive skills in their job posts or preliminary job screenings due to the inherent subjectivity of these traits, and may instead choose to screen for non-cognitive skills during subsequent candidate interviews. However, although assessing non-cognitive skills is often a challenge for employers, properly signaling both the expected non-cognitive and cognitive skills in job advertisements is an important step in establishing clear employee expectations.

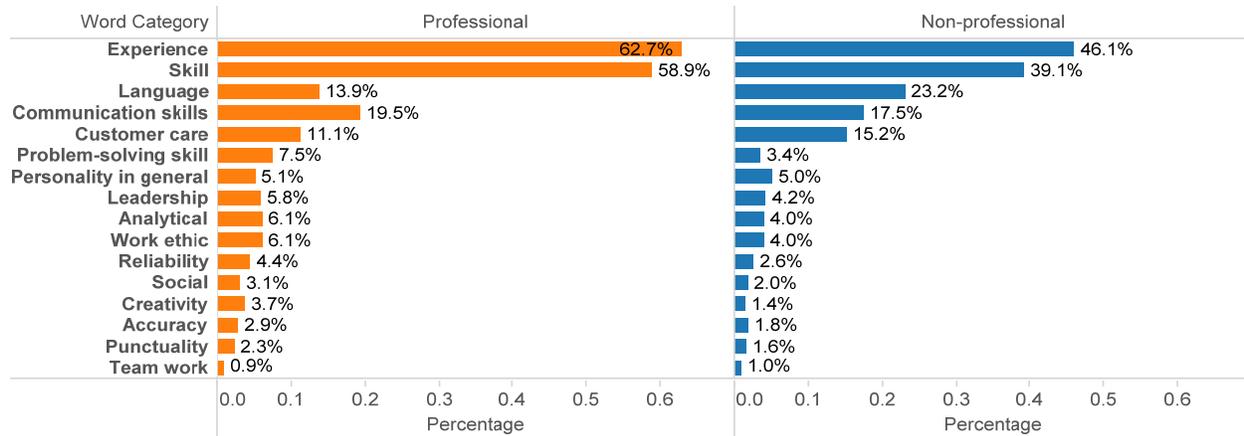
**Figure 5: Text Clouds for Job Descriptions by Occupational Level, 2014**



Source: Authors’ analysis using Babajob data.

Note: Text clouds are visual representations of the most frequently used words in job descriptions. The size of the word corresponds with the number of times it was used. Of the 186,829 job advertisements published on Babajob in 2014, 32,665 had job descriptions, 7,396 of which were for professional-level positions and 25,269 of which were for non-professional positions.

**Figure 6: Percentage of Job Advertisements Using Keyword Clusters by Occupational Level, 2014**



Source: Authors' calculations using Babajob data.

The use of text analysis can provide far richer information on the employee characteristics demanded by employers than can traditional surveys. By associating key cognitive, non-cognitive and technical skills with different occupations, the data can inform job seekers of the skills they should strive to acquire in order to become more competitive in the job market. The text analysis can be extended if shortlisted job seekers possess the demanded skill sets, by analyzing their CVs. These analyses can also provide important indicators to policy makers and educational institutions regarding the types of skills training that should be prioritized.

### 5.3. Observing Job-Search Behavior and Improving Skills Matching

Behavioral economics has emerged as an important tool of international development research, as it enables a more comprehensive analysis of the psychological, social and cultural factors that affect decision making and influence social and economic behavior (World Bank, 2015). While development economists are attempting to collect more detailed information on the personality traits and psychological characteristics of individuals, the methodological limitations of surveys, questionnaires and other traditional techniques, which often rely on self-reporting of behavioral or personality information, can make it difficult to capture behavioral data. As a result, these data are often collected through experimental research methods.

The extensive transactional information saved on online job portals offers a unique source of observed behavioral information on job-seeking and recruiting trends. Job seekers face a variety of constraints when applying for a job, such as their urgency to find work, the availability of jobs in their chosen field, as well as salary and location preferences. These diverse constraints result in substantial differences in job-seeking behavior, which standard sample surveys cannot capture. While the literature on the use of online job-portal data in behavioral analysis is currently very limited, the application of behavioral analysis to non-

experimental data could open up new avenues for reducing frictional mismatches between the supply and demand for workforce skills. Behavioral-economics analysis could help policy makers better understand job seekers' motivations and more effectively mitigate skills mismatches in the labor market. The analysis presented below examines two aspects of job-seeking strategies that can be used to guide policy making in India: (i) the number of job categories that applicants apply to and (ii) the timing of their applications.

### ***Application of the Analysis to Babajob Data***

Job seekers may not necessarily apply for jobs that utilize the skill sets that they have acquired through their education or training. Employer demand and the personal circumstances of the individual can both influence a job seeker's behavior. Those facing chronic unemployment or a serious financial constraint may adopt various coping strategies (Lazarus 1991; Lazarus and Folkman 1984; Liu, Huang and Wang 2014). For example, educated job seekers may choose to take lower-skilled jobs. However, their less-educated counterparts often face more limited employment prospects.

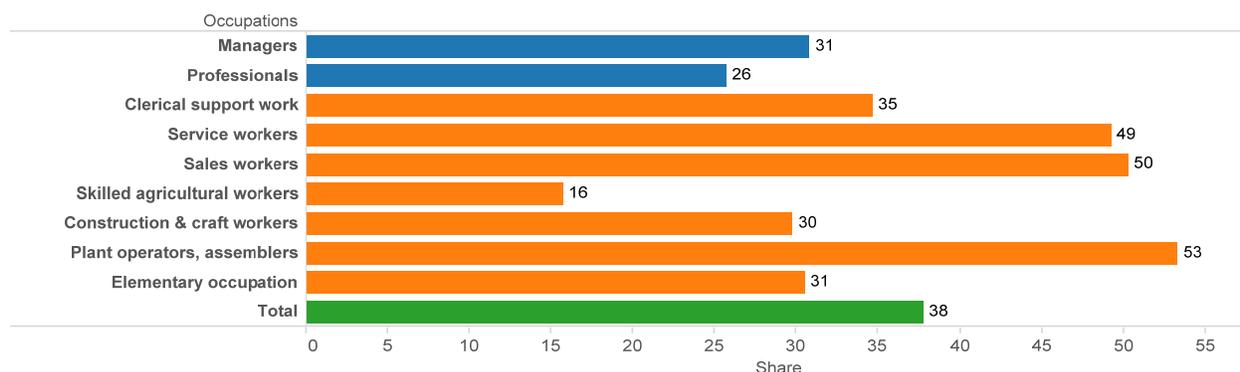
Figure 7 shows the proportion of job seekers who applied to only one occupational category on Babajob. Job seekers looking for managerial and professional-level positions were generally more likely to apply for jobs across multiple occupational categories.<sup>10</sup> Only 31 percent of applicants for managerial positions applied exclusively for jobs in one occupational category, while 26 percent of applicants for professional positions focused on a single category. This suggests that more-skilled job seekers have a greater degree of employment flexibility and are more willing to apply for diverse jobs. Among non-professional job seekers, some types of workers tended to restrict themselves to one specific job category, while others applied to jobs across various categories. About 50 percent of service workers, sales workers and plant operators and assemblers applied exclusively to jobs in one occupational category. This may indicate that workers in these categories are attempting to leverage a particular skill set or previous work experience. For example, 52 percent of beauticians and 53 percent of drivers applied to only one occupational category. However, skilled agricultural workers, workers in elementary occupations,<sup>11</sup> and construction and craft workers were more likely to apply to multiple occupation categories. Workers in these categories may have less-specialized skills and could be more sensitive to wages, employment terms or other factors besides job type. Clerical-support applicants tended to have an especially high degree of flexibility in terms of the employment categories to which they applied.

---

<sup>10</sup> This refers to the 26 occupational categories used by Babajob.

<sup>11</sup> Elementary occupations consist of simple and routine tasks which mainly require the use of hand-held tools and often some physical effort.

**Figure 7: Share of Workers Applying for Jobs in Only One Occupational Category by Main Occupation Choice**



Source: Authors' calculations using Babajob data.

Note: The analysis was conducted by excluding job seekers who applied for multiple jobs. Of 90,258 job seekers who submitted at least one job applications between January and December 2015, 88,098 (99%) submitted multiple applications.

The analysis also examined the relationship between the behaviors of job posting and application. Yamauchi et al. (forthcoming) applied time-series analysis to Babajob data to understand the relationship between the demand for and supply of workforce skills. The numbers of advertisements per firm and applications per job seeker were used to represent the demand and supply sides, respectively, and both reflect trends on the intensive margin. Focusing on business-process outsourcing (BPO) jobs advertised on a daily basis between 2012 and 2015 in Bangalore, the analysis found that job seekers responded quickly to job advertisements, suggesting that the supply of required skills, at least in terms of quantity, was not a binding constraint, and that the quantitative skills gap between the demand and supply for BPO jobs in Bangalore appears to be relatively small. Moreover, an increase in the number of job advertisements was found to have a surprisingly fast and significantly positive impact on the number of job applications received, with most job seekers applying for a posted position within one to two days.<sup>12</sup> The above results suggest that (i) job advertisement information circulates very quickly through the Babajob job portal and (ii) the supply of required skills is not a constraint for BPO jobs.<sup>13</sup>

Such job seekers' responsive behavior may reflect the success rate of their job search activity. The analysis also examined employer preferences regarding the timeframe for completing the hiring process. From employers' point of view, while they wish to recruit the candidate whose qualifications and skill set best matches their selection criteria, they cannot search for this candidate indefinitely. As a result, employers

<sup>12</sup> The demand for skills is not sensitive to application behavior primarily because of Babajob's sequential system structure. Both series have serial correlations.

<sup>13</sup> Other job categories show different patterns, and by comparing different job categories Yamauchi et al. (forthcoming) provide deeper insight into the supply and demand for workforce skills.

will attempt to strike a balance between finding the best candidate and managing the cost of the job search (Simon 1955; Iyengar, Wells, and Schwartz 2006).

An analysis of recruitment behavior among BPO employers on Babajob shows that the timing of application submissions affects the likelihood of being shortlisted (Figure 8). Of the 173,044 applications submitted in response to the 919 BPO jobs that were posted between January and October 2015 in the 10 cities<sup>14</sup> as part of the shortlisting service package,<sup>15</sup> 27 percent were submitted within 5 days of the posting, 16 percent were submitted between 6 and 10 days, and 76 percent were submitted within one month of the posting. Overall, 29.9 percent of applicants were shortlisted.

An econometric analysis using a probit model<sup>16</sup> shows a clear and statistically significant advantage to applying earlier in the recruitment process. Applications were grouped into five categories based on how long after the job posting they were submitted: within 120 hours (5 days), within 240 hours (10 days); within 480 hours (20 days), within 720 hours (30 days) and after more than 720 hours. Controlling for location, number of vacancies, month, and applicants' basic characteristics, applications submitted within 5 days were 15.6 percentage points more likely to be shortlisted than applications submitted after 30 days. This advantage narrows to 9.9 percentage point for applications submitted within 10 days, 2.2 percentage point for applications submitted within 20 days and 1.4 percentage points for applications submitted within 30 days. This suggests that employers are striving to fill positions quickly and are more likely to focus on applications submitted within 10 days of a job posting. It also suggests that job seekers would maximize their probability of being shortlisted by applying as quickly as possible. One prospective area for further research would be to quantify the costs involved in the trade-off between filling a position quickly and finding a more qualified candidate.

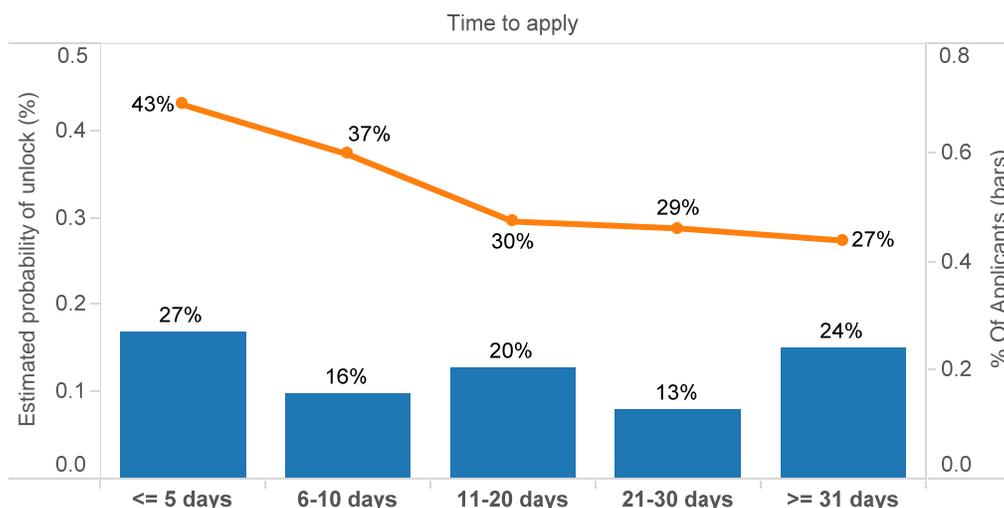
---

<sup>14</sup> These are the 10 cities with the largest number of job advertisements on Babajob. Advertisements were open for standard 90 days.

<sup>15</sup> The process of shortlisting is captured through the “unlock” function in the Babajob system. It allows employers to obtain the contact information of job seekers.

<sup>16</sup> In economics, a probit model is a type of regression used to describe dichotomous or binary outcome variables.

**Figure 8: Estimated Probability of Being Shortlisted for BPO Jobs, by Amount of Time between the Job Posting and the Submission of the Application**



Source: Authors' analysis using Babajob database

Note: The figure shows the estimated probability of being shortlisted by using a probit model, controlling for location, number of vacancies, month, and applicants' basic characteristics.

#### 5.4. Predictive Analysis of Skills Demand

Online job-portal data can facilitate predictive analysis based on historical labor market trends and emerging dynamics. In many developing countries the institutions charged with building workforce skills often struggle to improve the employment prospects of their trainees due to skills mismatches in the labor market. While analysts and policy makers often attempt to forecast changes in skills demand in order to better target training programs, these efforts are frequently constrained by a lack of timely and accurate information on labor market conditions.

Workforce surveys may take months or years to publish, resulting in a substantial lag between when information is collected and when it is available to policy makers. Traditionally, many countries have used “manpower planning” approaches to forecast labor demand. Manpower planning focuses on headcount imbalances and uses highly subjective employer surveys to determine the number of workers required in each occupation.<sup>17</sup> The limitations of this approach have increasingly led policy makers to replace it with different types of labor market analysis (Psacharopoulos 1991).

Some international organizations conduct workforce surveys and publish information on employment trends at the regional or country level (e.g. CEDEFOP 2010 and ILO 2016). Econometric modeling can be

<sup>17</sup> See Spalletti (2008) for a review of manpower planning and forecasting approaches.

used to incorporate macroeconomic and demographic information and address data constraints, such as missing data from non-survey years (ILO 2010). However, labor market forecasts by occupation, training type or educational attainment level are not standard, primarily due to a lack of data and difficulties in consistently integrating data into economic projections (Maier, Monnig and Zika 2015).<sup>18</sup>

In this context, the use of online data for “nowcasting”<sup>19</sup> and forecasting offers new ways to analyze labor market trends and skills demand, which can largely overcome the subjectivity and time-lag issues associated with workforce surveys. For example, Askitas and Zimmermann (2009) used Google keyword searches to conduct a predictive analysis of unemployment rates in Germany. In the wake of the 2008 global economic crisis, this innovative method for using online data proved to be a valuable tool to predict economic behavior, and strong correlations were discovered between keyword searches and unemployment rates. More recently, Vicente, Lopez-Menendez and Perez (2015) examined the possibility of using internet search data and business-confidence indicators to predict unemployment trends in Spain.

### ***Application of the Analysis to Babajob Data***

Online job-portal data can provide a continuous source of up-to-date, objective information on the demand for and supply of skills in different industries, occupations or locations. A simple predictive analysis using Babajob data has the potential to forecast trends in three variables, including the demand for skills as listed in job postings, the supply of skills as reported in job seeker profiles, and the wages being offered for new positions. However, without the presence of relevant and timely external data for cross-checking the composition of skills demand and supply against, nowcasting or forecasting the skills demand and supply in quantity is constrained.<sup>20</sup> The analysis of wage offers is the most robust among the three variables, considering that the wage offers posted on Babajob are influenced by the labor market for its competitiveness.

### ***Analyzing Wage Trends and Forecasting***

Areias et al. (forthcoming) analyzed the patterns of wage growth and distribution across different locations using 50,000 job advertisements posted in 20 cities with the largest number of advertisements. Wages were deflated using state-level urban consumer price index (CPI) obtained from the Reserve Bank of India.

---

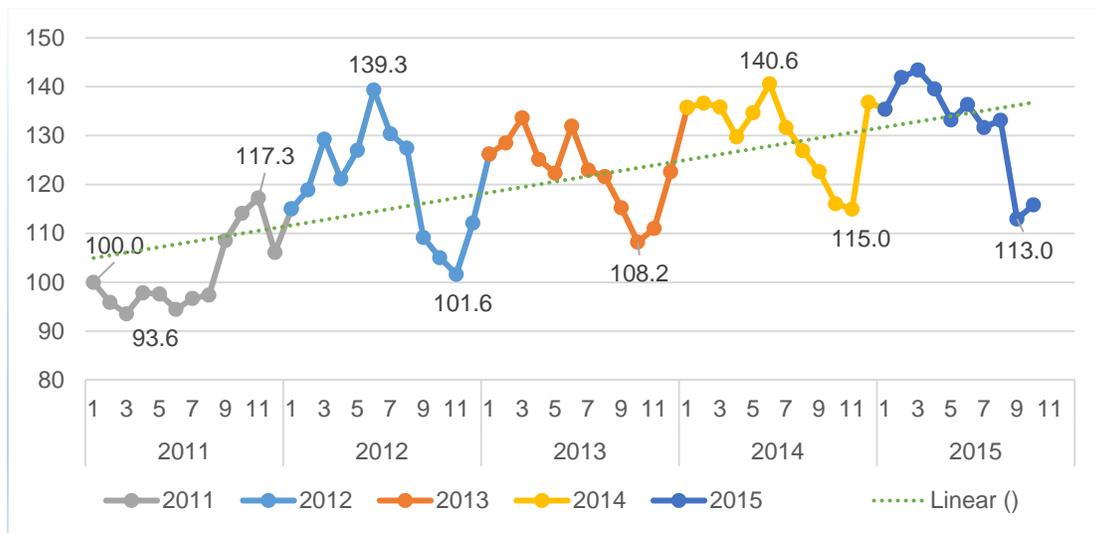
<sup>18</sup> Econometric models can be used to forecast economy-wide employment. See, e.g., ILO (2013).

<sup>19</sup> “Nowcasting” is used to estimate economic conditions in the present.

<sup>20</sup> Due to absence of national labor statistics since 2012 at the time of conducting analysis, it was not possible to compare post 2012-period of Babajob data (which this paper mostly uses) with the growth trends of national skills demand and supply.

Controlling for characteristics of jobs, including occupational compositions, locations, type of contract and shifts, requirement of qualifications such as years of experience and gender specification, and characteristics of the employing firms, the analysis revealed historical growth trends by occupation and location as well as a unique seasonal patterns of fluctuations in wage offers. Figure 9 presents the stylized trend of real wage offers for all jobs in 20 cities. Taking the index value of 100 for the real wage of January 2011, the wage offers grew steadily over time. However, the wage offers are subject to a consistent pattern of seasonal fluctuations every year. The wage offers tend to grow towards summer months, such as July and August, and tend to face a sudden fall around in October and November. The trend of wage offers is considered to reflect the demand and supply of skills. Considering the fact that Babajob provides many entry-level job opportunities for fresh graduates, this annual cycle shows a correlational pattern with the timing students' graduation and of job hunting, as well as with changes in skills demand associated with seasonal events.

**Figure 9: Stylized Trend of Real Wage Offer Index (January 2011=100), 2011-2015**



Source: Authors' analysis using Babajob database

Note: The figure shows the estimated real wage trend after controlling for variables of occupation and type of contracts, qualification requirements, and employers' types. It includes 20 cities in India.

By extrapolating from current dynamics, analysts and policy makers can use job-portal data to predict future trends in the labor market. The econometric model developed by Areias et al. (forthcoming) can be used to predict the real wage offers by location and occupation. Such predictions, including both nowcasting and forecasting of the near future, would provide valuable information to policy makers to understand the

growth of skills demand.<sup>21</sup> Moreover, the model can easily be updated with new data to deliver timely analysis of emerging labor market dynamics.

Prospective job seekers and training institutions can use these forecasts to tailor their education and training curricula to reflect trends in the demand for specific skills. Similarly, policy makers can identify skills gaps and make demand-driven policy decisions. These gaps may reflect the types of skills demanded or differences in demand conditions across geographical locations. Real-time data can also provide an early warning of disruptions in the labor market. For example, slowing wage growth during 2015 may pose a sign of weakening trend in workforce skills demand.

### **5.5. Experimental Studies**

Firms often pilot new business strategies to test their effectiveness. Horton and Tambe (2015) found that experimentation in digital markets is often simple and inexpensive, since interactions between users are computer-mediated and interfaces are relatively easy to modify. Horton (2016) conducted an experiment in which oDesk<sup>22</sup> introduced algorithmic recommendations for employers about which workers to hire. He found that the employers with technical job vacancies that received recruiting recommendations had a 20 percent higher fill rate compared to the control group, and there was no evidence that the treatment crowded out the hiring of non-recommended candidates. Gee (2015) collaborated with LinkedIn on an experiment that tested whether access to additional information affects job application behavior and found that additional information increased completed job applications by 1.9 percent overall, and by as much as 6 percent among female applicants.

#### ***Application of the Analysis to Babajob Data***

Babajob is currently considering an experimental study designed to improve employment transitions and promote better matches between job seekers and employers.<sup>23</sup> Frictional unemployment is the time that workers spend between jobs, which can be extended if inadequate information constrains employers' capability to select good candidates. This is particularly the case for subjective employee characteristics and skills that are not easily observed. Information on prospective employees is costly to acquire (Stigler 1962), and under certain circumstances imperfect information can create suboptimal equilibria (Spence, 1974; Fang and Moro, 2011). Facilitating the flow of information between workers and employers can

---

<sup>21</sup> Forecasting changes in workforce skills demand and in the supply of workers would be of great interest to both employers and job seekers. However, due to absence of external data to validate against at the time of data analysis, such an analysis is not conducted.

<sup>22</sup> oDesk, which is now renamed as Upwork, is a job portal for freelancers.

<sup>23</sup> This study is being funded by the Strategic Research Program and the Jobs Umbrella Trust Fund.

generate net gains in the labor market. Grosh et al. (2015) conducted a randomized experiment involving a labor market matching service in Jordan. This experiment aimed to reduce search costs by offering job-matching services based on educational backgrounds and psychometric assessments. However, the study concluded that what appeared to be frictional unemployment was in fact the result of the explicit preferences of job seekers, 28 percent of whom turned down the opportunity for an interview; among those who received job offers, 83 percent rejected the offer or quit the job shortly after taking it.

The situation in India may be different, as unemployment is often attributed to an insufficient number of job opportunities. A set of randomized control trials is currently being designed in an effort to reduce information asymmetry between job seekers and employers. The trials will focus on the role of non-cognitive skills among job seekers. In an earlier study, Indian employers hiring recent engineering graduates rated behavioral skills such as teamwork, reliability, leadership and willingness to learn, creative thinking and problem-solving skills, and specific knowledge and technical skills as important qualifications. Two out of three of these employers reported that most of these skills were “very” important, but that they were only “somewhat” satisfied with graduates’ skills levels (Blom and Saeki 2011).

Though reducing the frictional unemployment caused by limited information is an important economic objective, there is not yet sufficient evidence to inform effective policies. In the area of non-cognitive skills, a growing number of empirical studies have incorporated “big-five traits”, a psychological instrument for the classification of personality traits, to explain employment probabilities and wages in the US and European countries,<sup>24</sup> contributing to the international knowledge of how non-cognitive traits are correlated with labor market outcomes. The Skills Toward Employment and Productivity (STEP) surveys, initiated by the World Bank, also use the big-five-based instrument to capture personality traits (Pierre et al. 2014). While these studies assessed the contribution of non-cognitive skills to employment outcomes and wages, information on how employers view non-cognitive skills is scarce. The proposed experimental study adds to the growing literature on non-cognitive skills and labor market outcomes by supplying new knowledge on how the availability of non-cognitive skills information contributes to employment outcomes.

Further studies should strive to identify the types of information employers are attempting to acquire about potential hires and the extent to which skills-matching improves when information gaps are narrowed. Specific attention should be paid to employers’ demand for information about the non-cognitive skills of

---

<sup>24</sup>See, e.g., Heineck and Anger (2010); Mueller and Plug (2006); Nyhus and Pons (2005); and Wichert and Pohlmeier (2010). See Almlund et al. (2011) for a comprehensive review of the literature.

potential hires, and how such information contributes to job matching. Babajob data can contribute to an analysis of the role of information on non-cognitive skills in the job-search process.

## **6. Additional Uses of Online Job-Portal Data and Job-Search Platforms**

Beyond the five potential applications described above, there are a number of additional possibilities for using big data to improve labor market outcomes. The following section describes other prospective avenues for using online job-portal data and job-search platforms to improve skills-matching and generate more policy-relevant information.

### *Understanding the Emerging Skills Requirements of New Technologies*

Rapid technological change is driving the evolution of business practices and shaping the demand for workforce skills. New technologies often create new job types that do not fit into the existing scheme of employment categories and workforce skills. In this dynamic environment, analyzing trends in job titles can help to identify emerging skills requirements.

The US Department of Labor has developed an Occupational Information Network (O\*NET), which comprises a regularly updated database of occupational characteristics and employee information.<sup>25</sup> It describes the knowledge, skills and abilities required for different positions, as well as the specific tasks and activities involved in each job type. O\*NET collects this information by randomly sampling businesses and workers in each job category. Online job-portal data can complement this type of database by providing real-time information on new jobs, employee requirements and scopes of work. In case of India, it is useful to apply this analysis to review the skills classification and fit them against the National Skills Qualification Framework<sup>26</sup> for improving the relevance of training.

Text analysis can be especially useful for linking emerging occupations with their required skill sets. Moreover, algorithms based on closely related skill descriptions can allow for more precise classifications of jobs with ambiguous titles. For example, “computer programmer” can be refined to “application developer,” “web programmer,” “web designer,” etc. based on the job description.

### *Advanced Matching Services for Employers and Job Seekers*

---

<sup>25</sup> See <http://www.onetcenter.org/> for more information.

<sup>26</sup> National Skills Qualification Framework (NSQF) is a competency-based framework that organizes all qualifications according to a series of levels of knowledge, skills and aptitude.

More advanced forms of text analysis can use machine-learning techniques to improve skills matching and enable workers to seek training in high-demand skills. By creating algorithms to alert suitable candidates to newly posted jobs, online job portals can reduce search costs for both employers and job seekers. Algorithms can also help employers describe in greater detail both the technical skills and non-cognitive skills required for the position, and portals can identify suitable job seekers based on characteristics that go beyond observable qualifications such as academic degrees and years of experience. This type of content-based matching can also be useful to job seekers, especially those whose interests and qualifications are not restricted to a single occupational category.

Collecting additional information on non-cognitive skills could improve the quality of job seeker information. Employers typically rely on interviews to assess non-cognitive skills, which tend to be abstract and subjective. Encouraging employers to identify the type of non-cognitive skills that they require and then assessing the non-cognitive skills offered by job seekers could reduce employers' reliance on time-consuming interview processes. LinkedIn and other job-matching services are already pursuing similar initiatives and have achieved some preliminary success (e.g., Gee, 2015; Horton, 2016).

### ***Limitations and Caveats***

There are a number of important limitations and caveats to the use of big data in general, and online job-search portal data in particular. These include compliance and privacy concerns, sample representativeness and selection biases, and overall data quality. Askitas and Zimmermann (2015) identify privacy issues and data ownership as major challenges involved in big data analytics. While data collected by private firms offer key insights that can inform better public policies, the legal and ethical frameworks for obtaining and using such data are not always clearly established. Moreover, big data analytics often involve micro-level data associated with individual transactions, which can give rise to serious privacy concerns.

The representativeness of the data is also a major issue, as online job-search platforms are subject to self-selection biases and other forms of sampling error. Carnevale et al. (2014) argued that online job-portal data are vulnerable to systematic errors arising from how employers and job seekers use them. For example, online job advertisements tend to be skewed toward upper-level and professional positions, while lower-level and unskilled positions are often filled through other channels. Babajob was selected for this study due in part to the fact that its postings tend to include more entry-level and informal jobs than other job-search portals. In addition, while barriers to internet access can restrict the type of job seekers who apply for online positions, Babajob also provides phone-based access to job seekers, eliminating the necessity of

internet access. However, the voluntary nature of participation in an online job-search portal creates an inevitable degree of self-selection bias.

There have been various attempts to correct representation issues, including through use of statistical weights and techniques for predicting missing data (Kureková, Beblavý, and Thum-Thyssen 2015). In the case of Babajob, external data can be used as a statistical benchmark to normalize biases, but the lack of sufficiently frequent external data to describe the contemporary Indian labor market remains a constraint. Unstandardized data quality is also a potential source of statistical error, as the data are usually not collected under a specific survey framework and may therefore suffer from an unusually large amount of missing information or internal inconsistencies. Job seeker information is also self-reported and may be inaccurate. However, measurement errors and sampling biases are a risk in most forms of statistical research, and if properly addressed they should not invalidate its results.

## **7. Conclusions**

Research using online job-portal data has enormous potential to inform public policy and advance the academic discourse on labor economics and workforce skills development. These data are not only more frequent and larger in volume than those obtained from traditional employer and workforce surveys, but they also present a greater variety of information and a more intense degree of granularity. Online job portals collect information on employers' actual demand for skills as expressed through their real business activities, as well as data on job seeker characteristics reported in a context with real economic consequences. Big data analytics has opened new avenues for objectively monitoring workforce skills demand with a wide array of applications for business practices and public policy. The analysis presented above yields five key conclusions.

First, governments should leverage online job-portal data to enhance their labor market policies. Many governments continue to rely on periodic sample-based employer and workforce surveys. However, traditional surveys are conducted too infrequently and their analytical lag time is too long to capture emerging trends in labor market dynamics and demand for workforce skills. While surveys will remain useful for analytical and policy purposes, big data analytics can support more informed decision making by providing real-time data that encompass variables which cannot be observed through any other statistical technique. Moreover, online job-portal data reflect real business activities, not abstract survey questions. These data can provide a richer understanding of employers' needs, constraints and transaction costs and uncover potential opportunities to address frictional unemployment and other labor market inefficiencies. Job-portal data can also help determine whether apparent skills gaps reflect a real deficit in the necessary

skills or an obstacle to effective job matching. In an era of rapidly changing technology, policy makers require up-to-the minute information on new job types and emerging trends in the demand for workforce skills. Online job-portal data can be used to facilitate demand-driven training, and policy makers should explore ways to link this information to skills development programs or job-placement services.

Second, policy makers can use big data analytics to develop a more detailed and timely understanding of the characteristics of the labor supply. Online job-portal data can provide information on the interests and qualifications of recent graduates, as well as the features of specific labor market participants, such as women, young people, or workers in the informal sector. Identifying the dynamics that drive labor force participation among women and young people, as well as their unique job-search constraints, could also contribute to greater labor productivity and support the achievement of key social development goals. Moreover, understanding the differences between formal and informal sector jobs and workers, how wages are determined, how job matching takes place, and what skills are in shortage or surplus in the informal sector are particularly crucial to the development of effective labor market policies in developing countries.

Third, online job-portal data are more valuable when used to complement other sources of information. As issues with representativeness are a major limitation on data quality, periodic surveys of the entire labor market are essential for normalizing samples and verifying data reliability. In addition, hybrid statistical techniques using both big data analytics and more traditional methodologies can reveal dimensions of the labor market that are otherwise unobservable.

Fourth, analytical results from online job portals and their policy implications should be made public, so long as privacy concerns can be effectively addressed. Ever-expanding access to internet technology and social networks is generating a wealth of information on human behavior, raising important questions about data confidentiality and ownership. Once an appropriate legal framework has been established and sound privacy controls are in place, information on current labor market trends, including job availability, wage growth and the demand for workforce skills, should be provided to the public in order to better inform personal and firm-level employment, education and training decisions. Students in particular can use online job-portal data to determine the skills that they may wish to acquire. Data visualization can assist in conveying complex information on employment dynamics to the public.

Fifth, the analytical results from online job portals can inform training programs to make them more demand-based. Training programs are often unable to catch up with the speed of changing labor market demands driven by fast technological change, and skills training institutions continue to provide outdated training with outdated training materials and curricula. The real-time big data analytics will provide the

trends in the skills demand and identifies shortages of skills. Such knowledge can be used to inform training institutions and contribute to adjusting skills development curriculum and improving job placement service centers for the improved relevance of training and matching support.

## References

- Aker, J.C., 2011. Dial “A” for agriculture: a review of information and communication technologies for agricultural extension in developing countries. *Agricultural Economics*. 42(6): 631-647.
- Almlund, M., A.L. Duckworth, J. Heckman, and T. Kautz, 2011, Personality Psychology and Economics. *Handbook of the Economics of Education* 4(1), 1-181.
- Areias, C.A., F. Yamauchi, S. Nomura, and S. Imaizumi. Forthcoming. Understanding Wage Dynamics through Job Advertisements: New Evidence from an Online Job Portal in India, Manuscript, The World Bank.
- Askatas, N. and K. Zimmermann. 2009. “Google Econometrics and Unemployment Forecasting.” *Applied Economics Quarterly*, 55(2): 107-120.
- Askatas, N. and Zimmermann, K.F., 2015. The internet as a data source for advancement in social sciences. *International Journal of Manpower*. 36(1): 2-12.
- Beard, T. R., Ford, G. S., Saba, R. P., & Seals, R. A., 2012, Internet Use and Job Search. *Telecommunications Policy*, 36(4), 260-273.
- Beblavý, M., B. Fabo, B. and K. Lenaerts. 2016a. “Skills Requirements for the 30 Most-Frequently Advertised Occupations in the United States: An Analysis Based on Online Vacancy Data.” CEPS Special Report. No. 132 March 2016.
- Beblavý, M., B. Fabo, B. and K. Lenaerts. 2016b. “The Importance of Foreign Language Skills in the Labour Markets of Central and Eastern Europe: An assessment based on data from online job portals.” *CEPS Special Report*, No. 129. January 2016.
- Blom, A. and H. Saeki. 2011. Employability and Skill Set of Newly Graduated Engineers in India. *Policy Research Working Paper 5640*. World Bank.
- Burga, P., I. Ritter, and M. E. G. Barreto, 2014, The effect of Internet and cell phones on employment and agricultural production in rural villages in Peru.
- Carnevale, A. P. T. Jayasundera, and D. Repnikov. 2014. Understanding online job ads data: A technical report. McCourt School on Public Policy, Center on Education and the Workforce, Georgetown University, Washington, D.C.
- CEDEFOP. 2010. Skills Supply and Demand in Europe. Medium-Term Forecast Up to 2020. European Centre for the Development of Vocational Training. Luxembourg, Publication Office of the European Union.

- Cunningham, W.V. and P. Villaseñor. 2016. “Employer voices, employer demands, and implications for public skills development policy connecting the labor and education sectors.” *The World Bank Research Observer*, 31(1): 102-134.
- Daroesman, R. and A. Weidemann. 1982. “The Role of Newspaper Advertisements in Employment Recruitment in Jakarta” *Bulletin of Indonesian Economic Studies*, 18(3), 116-120.
- Einav, L. and Levin, J.D., 2013. The data revolution and economic analysis (No. w19035). National Bureau of Economic Research.
- Fang, M. and A. Moro, 2011, Theories of Statistical Discrimination and Affirmative Action: A Survey, in *Handbook of Social Economics* Vol.1A, eds by J. Benhabib, A. Bisin and M.O. Jackson, 133-200.
- Gallivan, J. M. D. P. Truex and L. Kvasny. 2004. “Changing patterns in IT skills sets 1988-2003: A content analysis of classified advertising. *DATA BASE for advances in information systems*, 35: 64-87.
- Gee, L.K., 2015. The more you know: Information effects in job application rates by gender in a large field experiment. Available at SSRN 2612979.
- Groh, M., McKenzie, D.J., Shammout, N. and Vishwanath, T., 2014. Testing the importance of search frictions, matching, and reservation prestige through randomized experiments in Jordan. *World Bank Policy Research Working Paper*, (7030).
- Hanushek, E.A. and L. Woessmann, 2008, The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature* 46(3): 607–668.
- Hanushek, E.A. and L. Woessmann, 2011, The Economics of International Differences in Educational Achievement. *Handbook of the Economics of Education*. Vol. 3. North Holland.
- Heineck, G., & Anger, S. 2010. The Returns to Cognitive Abilities and Personality Traits in Germany. *Labour Economics*, 17(3), 535-546.
- Horton, J. J. 2016. The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment. Available at SSRN 2346486.
- Horton, J.J. and Tambe, P., 2015. Labor Economists Get Their Microscope: Big Data and Labor Market Analysis. *Big Data*, 3(3), pp.130-137.
- Huizen T. V. and J. Plantenga. 2014. “Job Search Behaviour and Time Preferences: Testing Exponential Versus Hyperbolic Discounting.” *De Economist*. 162: 223-245.
- (IFC) International Finance Corporation. 2014. *India Country Profile 2014*. The World Bank Group.

- (ILO) International Labour Office. 2010. Trends Econometric Models: A Review of the Methodology. ILO Employment Trends Unit.
- (ILO) International Labour Office. 2013. The Philippines Employment Projections Mode; Employment targeting and scenarios. Employment working paper No. 140. Employment Trends Unit.
- (ILO) International Labour Office. 2016. *World Employment Social Outlook: Trends 2016*. Geneva.
- Iyengar, S. S., R. E. Wells, and B. Schwartz. 2006. "Doing Better but Feeling Worse: Looking for the Best Job Undermines Satisfaction." *Psychological Science*. 17(2): 143-150.
- Klonner, S., & Nolen, P. J., 2010, Cell Phones and Rural Labor Markets: Evidence from South Africa. Proceedings of the German Development Economics Conference, Hannover 2010 56, Verein für Socialpolitik, Research Committee Development Economics.
- Kroft, K., & Pope, D. G., 2014, Does Online Search Crowd out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist. *Journal of Labor Economics*, 32(2), 259-303.
- Kuhn, P., & Mansour, H., 2014, Is Internet Job Search Still Ineffective? *Economic Journal*, 124(581), 1213-1233.
- Kurekova, L. M. and Z. Žilinčíková., 2015. "Low-Skilled Jobs and Student Jobs: Employers' Preferences in Slovakia and the Czech Republic." IZA Discussion Papers, No. 9145.
- Kureková, L.M., M. Beblavý, and A. Thum-Thysen. 2015. "Using online vacancies and web surveys to analyse the labour market: a methodological inquiry." *IZA Journal of Labor Economics*, 4(1): 1-20.
- Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. META Group Research Note, 6, p.70.
- Lazarus, R. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press.
- Lazarus, R., and Folkman, S. (1984). *Stress, appraisal, and coping*. New York, NY: Springer.
- Lenaerts, K., Beblavý, M. and Fabo, B., 2016. Prospects for utilisation of non-vacancy Internet data in labour market analysis—an overview. *IZA Journal of Labor Economics*. 5(1): 1.
- Liu, S., Huang, J. L., and Wang, M. 2014. Effectiveness of Job Search Interventions: A Meta-Analytic Review. *Psychological Bulletin*. 140(4): 1009-1041.
- Maier, T., Mönnig, A. and Zika, G., 2015. Labour demand in Germany by industrial sector, occupational field and qualification until 2025—model calculations using the iab/inforge model. *Economic Systems Research*, 27(1): 19-42.

- Marinescu, I.E. and R. P. Wolthoff. 2015. "Opening the black box of the matching function: The power of words." IZA Discussion Papers, No. 9071.
- McAfee, A., and E. Brynjolfsson. 2012. "Big Data: The management Revolution." *Harvard Business Review*. 90: 60-68.
- Mueller, G., & Plug, E., 2006, Estimating the Effect of Personality on Male and Female Earnings. *Industrial and Labor Relations Review*, 3-22.
- Nyhus, E. K., & Pons, E., 2005, The Effects of Personality on Earnings. *Journal of Economic Psychology*, 26(3), 363-384.
- OECD. 2012. *Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies*. OECD Publishing. <http://dx.doi.org/10.1787/9789264177338-en>
- Pierre, G., M. L. S. Puerta, A. Valerio, and T. Rajadel. 2014. *STEP Skills Measurement Surveys. Innovative Tools for Assessing Skills*. Social Protection and Labor Discussion Paper No. 1421. Washington, D.C.
- Psacharopoulos, G. 1991. "From Manpower Planning to Labour Market Analysis." *International Labour Review*. 130(4): 459-474.
- Reimbsbach-Kounatze, C. 2015. "The Proliferation of "Big Data" and Implications for Official Statistics and Statistical Agencies: A Preliminary Analysis", OECD Digital Economy Papers, No. 245, OECD Publishing.
- Rutkowski, J. 2010. "Demand for Skills in FYR of Macedonia." Technical Note. World Bank, Washington, D.C.
- Savage M. and R. Burrows. 2007. The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.
- Savage M. and R. Burrows. 2009. Some further reflections on the coming crisis of empirical sociology. *Sociology* 43(4): 762–772.
- Shahiri, H., & Osman, Z., 2014, Internet Job Search and Labor Market Outcome. *International Economic Journal*, 1-13.
- Simon, H. A. 1955. "A behavioral model of rational choice." *Quarterly Journal of Economics*. 59: 99-118.
- Snijders, C., Matzat, U., & Reips, U.-D. 2012. 'Big Data': Big gaps of knowledge in the field of Internet science. *International Journal of Internet Science*, 7, 1-5.
- Spalletti, S. 2008. The History of Manpower Forecasting in Modelling Labour Markets, Working Paper 18, Macerata, University of Macerata.

- Spence, M., 1974, Job Market Signaling. *Quarterly Journal of Economics* 87, 355–374.
- Stigler, George, 1962, Information in the labor market. *Journal of Political Economy* 70, no. 5:94–105.
- Tambe, P., 2014. Big data investment, skills, and firm value. *Management Science*. 60 (6): 1452-1469.
- Taylor, L., Schroeder, R. and Meyer, E., 2014. Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? *Big Data & Society*, 1(2),
- United Nations. 2012. Convention on the Elimination of All Forms of Discrimination Against Women. Consideration of reports submitted by States parties under article 18 of the convention. Combined fourth and fifth period reports of States parties. India.
- Vicente, M.R., López-Menéndez, A.J. and Pérez, R., 2015. Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting and Social Change*, 92: 132-139.
- Weichselbaumer, D. and Winter - Ebmer, R., 2005. A meta - analysis of the international gender wage gap. *Journal of Economic Surveys*, 19(3): 479-511.
- World Bank. 2010. *Stepping up Skills: For More Jobs and Higher Productivity*. World Bank Publications.
- World Bank. 2012. *More and Better Jobs in South Asia*. South Asia Development Matters. Washington, D.C.
- World Bank. 2015. *World Development Report 2015. Mind, Society, and Behavior*. Washington, D.C.
- World Bank. 2016. *World Development Report 2016. Digital Dividends*. Washington, D.C.
- Yamauchi, Futoshi, Ana Areias, Shinsaku Nomura and Saori Imaizumi, 2016, Power of Information Technology – A New Analysis of Job Seeking Behavior and Skill Distribution, Manuscript, The World Bank.