

Labor Market Analysis Using Big Data

The Case of a Pakistani Online Job Portal

Norihiko Matsuda

Tutan Ahmed

Shinsaku Nomura



WORLD BANK GROUP

Education Global Practice

November 2019

Abstract

Facing a youth bulge—a large influx of a young labor force—the Pakistani economy needs to create more jobs by taking advantage of this relatively well-educated young labor force. Yet, the educated young labor force suffers a higher unemployment rate, and there is a concern that the current education and training system in the country does not respond to skill demands in the private sector. This paper provides new descriptives about labor markets, particularly skill demand and supply, by using online job

portal data. The paper finds that although there is an excess supply of highly educated workers, certain industries, such as information and communications technology, lack workers who have specialized skills and experience. The analysis also finds that the exact match of qualifications and skills is important for employers. Job applicants who are underqualified or overqualified for job posts are less likely to be shortlisted than those whose qualifications exactly match job requirements.

This paper is a product of the Education Global Practice. 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://www.worldbank.org/prwp>. The authors may be contacted at 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.

Labor Market Analysis Using Big Data: The Case of a Pakistani Online Job Portal¹

Norihiko Matsuda
Tutan Ahmed
Shinsaku Nomura

JEL codes: C81, J01, J20

Keywords: Skill demand and supply, online job portal, big data, Pakistan

¹ This paper has been prepared as a background paper for an analytical project of the World Bank, *Pakistan: Skills Assessment for Economic Growth*. The authors are grateful to Naseeb Online Services (Private) Ltd., which operates the online job portal, Rozee.pk, for willingly collaborating and working together with the research team. Special thanks go to Monis Rahman for his support and interest in the collaboration and generous data sharing, and to Muhammad Khalid, Usman Haider, and Sameer Akbar Malik for operational task management and data and analytical work. The authors are grateful for helpful comments from Tazeen Fasih, Cristian Aedo, Cristina Isabel Panasco Santos, Melinda Good, and Rohini Pande. Maliha Hyder provided editorial support.

1. Introduction

Employment is a key challenge in Pakistan. While its unemployment rate is low at 6.0%,² the quality of employment is not high. Among the employed people, three-quarters work in informal sectors; 36% and 24% are self-employed and family workers, respectively; and only 12% are wage workers with a written contract (Pakistan Bureau of Statistics, 2015; Bossavie et al., 2018). The youth, despite being more educated, face a three times higher unemployment rate and are more likely to work in informal sectors than older people (Bossavie et al., 2018).³ Facing a youth bulge—a large influx of a young labor force—in coming years, the Pakistani economy needs to create more jobs by taking advantage of the relatively well-educated young labor force. However, the higher unemployment rate of the educated young labor force raises the concern that skill development in the current education and training system does not respond well to skill demands in industries.⁴

The objective of this paper is to provide new descriptives about labor market conditions and skill demand and supply in Pakistan based on novel data from a leading online job portal, Rozee.pk. Online job portal data have unique features that can supplement traditional data such as labor force surveys in four respects.⁵ First, online job portal data provide real-time information, whereas traditional data are published after a time lag of months to even a year. As technologies and economic conditions are rapidly changing, real-time information of labor markets is becoming more important. Second, the data provide rich, granular text information about skill demand and supply. For example, included in the data are the job's title, description, and qualification requirements and jobseeker's education, skills, professional experience.⁶ Third, the data include information about actual labor market transactions and the job matching process, which are invisible in traditional labor force and enterprise surveys. With this information, it becomes possible to analyze what affects, and can improve, job matching. Lastly, the data provide information linked between employers and jobseekers. The availability of this linked information allows labor market diagnosis to be well connected between the demand and supply sides, which is not easy with labor force survey and enterprise survey data that are separately conducted.

In addition to the novel features of the data, there are other reasons for analyzing online job portals. First, online job portals are increasingly common. A considerable amount of job postings in developed countries have been shifting from print media to websites (Kroft and Pope, 2014; Mang, 2012). Despite the fact that the internet access in developing countries is still limited (Shahiri and Osman, 2014), many people already use them.⁷ Second, online job portals can improve the efficiency of the job search and matching process. The literature finds that online

² A primary reason for the low unemployment rate is that most people in the country cannot afford to be not employed so engage in informal and self-employed jobs even though they are not necessarily satisfied with work conditions and compensations (Pakistan Bureau of Statistics, 2015).

³ The youth are defined as people aged 15 to 24 years.

⁴ Bossavie et al. (2018) find that the unemployment rate is 16% among those with postsecondary education but virtually zero among those with no and primary education.

⁵ For reviews of advantages and disadvantages of using online labor market data, see Einav and Levin (2014) and Kureková et al. (2015).

⁶ For example, Deming and Kahn (2018) analyze skill demands by examining text information in online data.

⁷ For example, the use of an online job portal is already common in India (Nomura et al., 2017).

job postings reduce information costs and unemployment rates and improve job match quality in developed countries (Beard et al., 2012; Kuhn and Mansour, 2014; Mang, 2012). Third, online job portal data, particularly the data from Rozee.pk, are relevant to specific challenges in Pakistani labor markets. The youth, especially well-educated young people, suffer a higher unemployment rate, and wage employment needs to increase. The country faces a gender disparity in labor force participation (LFP).⁸ The data from Rozee.pk are suitable to analyze these challenges because the online job portal is mostly used by well-educated young people for wage jobs, and many job postings are gender earmarked.

2. Online job portal data from Rozee.pk

Rozee.pk, founded in 2007, is a leading online job portal in Pakistan. This paper analyzes a data set consisting of 5.0 million jobseekers and 108,000 registered employers corresponding to 412,000 jobs posted on the platform. The data provided by Rozee.pk to the study team consist of four types of data sets: jobseekers, employers, job postings, and transactions.⁹ The data set of jobseekers is based on their resumes created at Rozee.pk and includes rich information about demography, education, professional experience, skills possessed, and current and desired salaries. The employer data set is based on employers' profiles at Rozee.pk. The job posting data set is an archive of all job postings and has virtually all information included in job postings, such as job titles, job descriptions, qualifications, and salary ranges. The transaction data set is a record of all applications as to who applied for which job postings on what date and time, and whose applications were viewed by employers. Information about application results, such as who was shortlisted, is also available, although this information is incomplete since application results are voluntarily updated by employers. This paper mainly uses the data period from 2012 to 2019 for the analyses, unless otherwise specified.

2.1 Description of the labor market domain represented by the data

One of the limitations of using online job portal data is that the data may not be representative of the entire labor market.¹⁰ Geographically, the jobseekers, employers, and job postings on Rozee.pk are concentrated in megacities, as illustrated in Figure 1. We categorize Lahore, Karachi, and Islamabad/Rawalpindi as first-tier cities; and Faisalabad, Gujranwala, Hyderabad, Multan, Peshawar, and Quetta as second-tier cities. Concentrations are observed in first-tier cities—Lahore, Karachi, and Islamabad/Rawalpindi—where 48% of the jobseekers, 79% of the employers, and 87% of the job postings are found. The second-tier cities—Faisalabad, Gujranwala, Hyderabad, Multan, Peshawar, and Quetta—constitute only 12% of the jobseekers, 8% of the employers, and 11% of the job postings. Many jobseekers in small cities also use the portal, as the proportion of jobseekers in the other domestic areas is 24%, whereas the proportion of job postings there is only 9%. This off-balanced ratio between jobseekers and job postings in the other domestic areas may mean that there are not enough job opportunities in these areas and

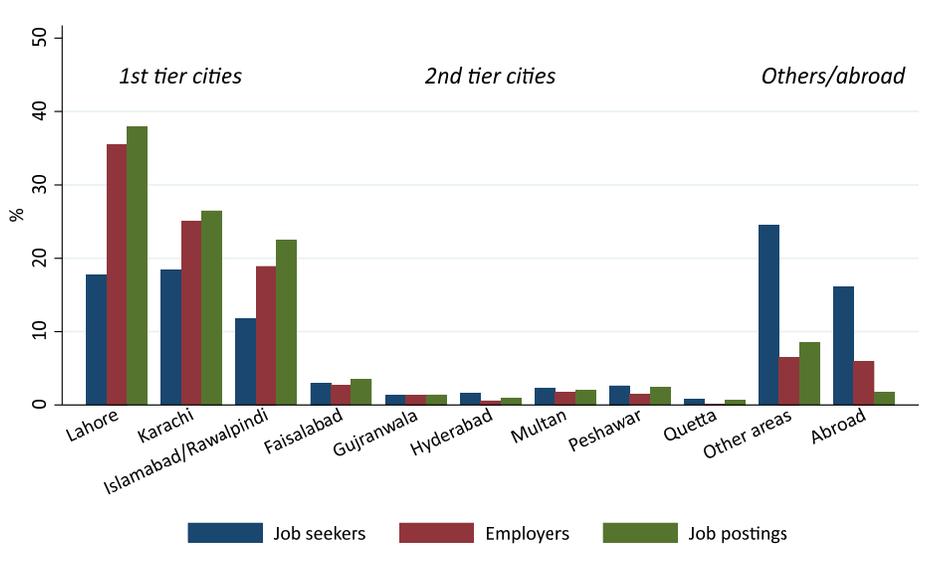
⁸ Only one in four women participate in the labor force while 82% of men do so (Amir et al.; 2018).

⁹ The data shared by Rozee.pk for this research were completely anonymized for the protection of the user privacy.

¹⁰ See Carnevale et al. (2014), Kureková et al. (2015) and Nomura et al. (2017).

therefore, jobseekers plan to migrate to megacities. Lastly, 1.7% of the job postings are in foreign countries.¹¹

Figure 1. Geographical distribution of jobseekers, employers, and job postings



Source: Authors' calculation using the data from Rozee.pk.

Note: (1) We refer to Lahore, Karachi, and Islamabad/Rawalpindi as first-tier cities and Faisalabad, Gujranwala, Hyderabad, Multan, Peshawar, and Quetta as second-tier cities. (2) Job locations are multiple for a substantial number of job postings. Thus, the sum of the percentages for job postings is greater than 100%.

Comparing Rozee.pk users and the entire labor market participants based on the Labor Force Survey 2014–15 and Enterprise Survey 2013 shows notable differences between them. First, the proportion of female workers is smaller in the online job portal than the entire labor market. Women constitute 21% of jobseekers in the portal but 25% of the entire labor force (Table 1). This may be because many female workers are agricultural and household workers, who mostly do not use Rozee.pk. Second, jobseekers in the online job portal are much younger (mean age is 26) than the entire labor force in Pakistan (mean age is 35). In the portal, 83% of all the jobseekers are aged 30 and below while only 44% of the entire labor force in the economy is aged 30 and below. The age distribution of the online job market is concentrated between 20 and 30 years old: the proportion of the 20- to 30-year-old labor force is 77% in the online job market but only 33% in the entire labor market. The composition of jobseekers by education in Rozee.pk is significantly different from the entire labor market of Pakistan. Of jobseekers at Rozee.pk, 86% have postsecondary degrees, and 32% have master's degrees or a Ph.D. By contrast, among the entire labor force, only 9% have a postsecondary degree.

¹¹ Since foreign jobs do not reflect labor markets in Pakistan, we exclude job postings in foreign countries in the analysis.

Table 1. Summary statistics of workers and employers in Rozee.pk and the entire labor market

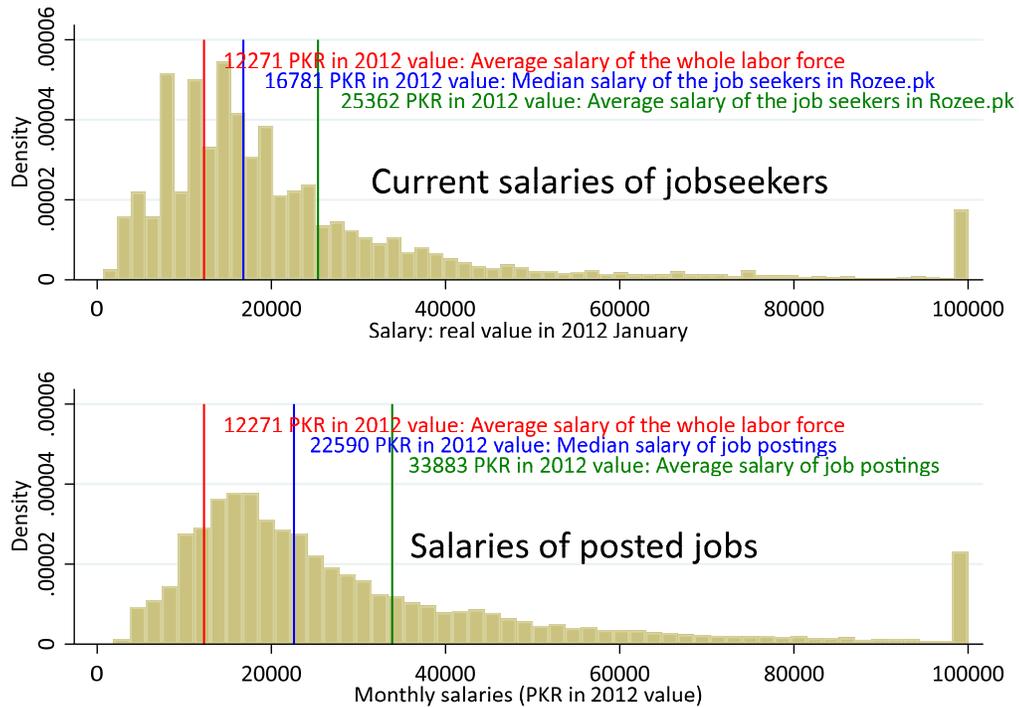
	Data from Rozee.pk		Labor Force Survey 2014-15, Enterprise Survey 2013	
	Mean	Std	Mean	Std
Workers				
Female	0.21	0.40	0.25	
Age	26.02	6.46	34.91	12.89
Education				
Secondary	0.10	0.30	0.32	0.46
Postsecondary	0.86	0.35	0.09	0.29
Salary (2012 PKR)	25,362	31,389	12,271	
Employers				
Size (# employees)				
Micro: 1-10	0.48	0.50	0.28	0.45
Small: 11-50	0.34	0.47	0.45	0.50
Medium: 51-100	0.09	0.28	0.09	0.29
Large: 101-600	0.07	0.25	0.13	0.34
Mega: 601+	0.03	0.18	0.05	0.21
Business age (years)	9.72	15.26	18.70	11.30

Source: Authors' calculation using Rozee.pk data, Enterprise Survey 2013, and Labor Force Survey 2014–15, except for the female proportion and salary in the far right columns. They are based on Pakistan Bureau of Statistics (2015).

Wages are very different between the Rozee.pk and the entire labor market. The average and median annual salaries of the labor force participants in Rozee.pk are PKR 25,362 and PKR 16,781, respectively, while the national average in 2014/15 is PKR 12,271 (Figure 2).¹² This difference may be in part because of the higher education of jobseekers in Rozee.pk. On the demand side, employers post higher paying jobs at Rozee.pk. According to the distribution of monthly salaries advertised in job postings (Figure 2), the average and mean advertised salaries are PKR 33,883 and PKR 22,590, respectively. These are much higher than the national mean salary, PKR 12,271.

¹² At the website of Rozee.pk, job seekers have an option to enter the information of current and desired salaries. Of job seekers, 45% enter the information of current salary, and 55% enter the information of desired salary.

Figure 2. Salary distributions of jobseekers and posted jobs



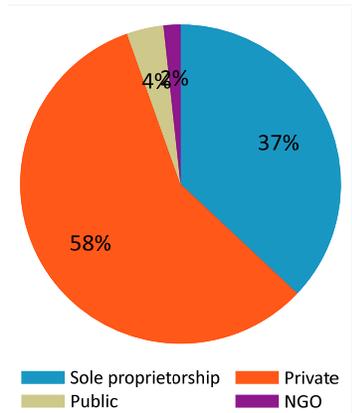
Source: Authors' calculation using the data from Rozee.pk.

Note: (1) Salaries are truncated from above at PKR 100,000 in this figure for a presentation purpose. When the average salary is calculated, salaries are not truncated. (2) The average salary of the whole labor force is based on Pakistan Bureau of Statistics (2015).

Smaller and younger businesses are more likely to use the online job portal. Figures 3, 4, and 5 illustrate employers' distributions by their ownership, number of employees, and business's age. Sole proprietorship (37%) and private businesses (58%) are dominant while public businesses (4%) and NGOs (2%) are very few. Regarding the business size in terms of the number of employees, the majority are micro (48%) and small businesses (34%). This proportion of micro businesses is higher than the proportion among the entire nonagricultural industries in Pakistan (Figure A1.1).¹³ The percentage of the businesses that started operations less than five years ago is 50% in the online job portal but only 11% in the entire nonagricultural industries in Pakistan.

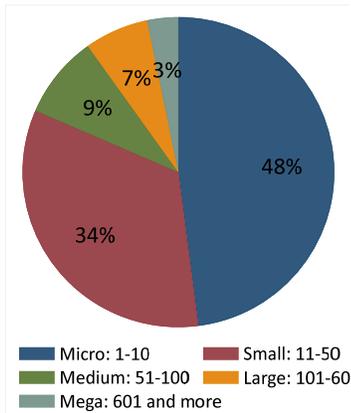
¹³ The statistics of the entire nonagricultural industries are based on the Enterprise Survey 2013. For the details of the survey, see World Bank (2015).

Figure 3. Firm ownership



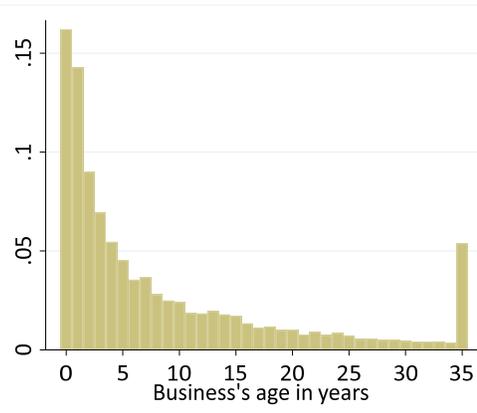
Source: Authors' calculation.

Figure 4. Firm size (number of employees)



Source: Authors' calculation.

Figure 5. Distribution of firm age

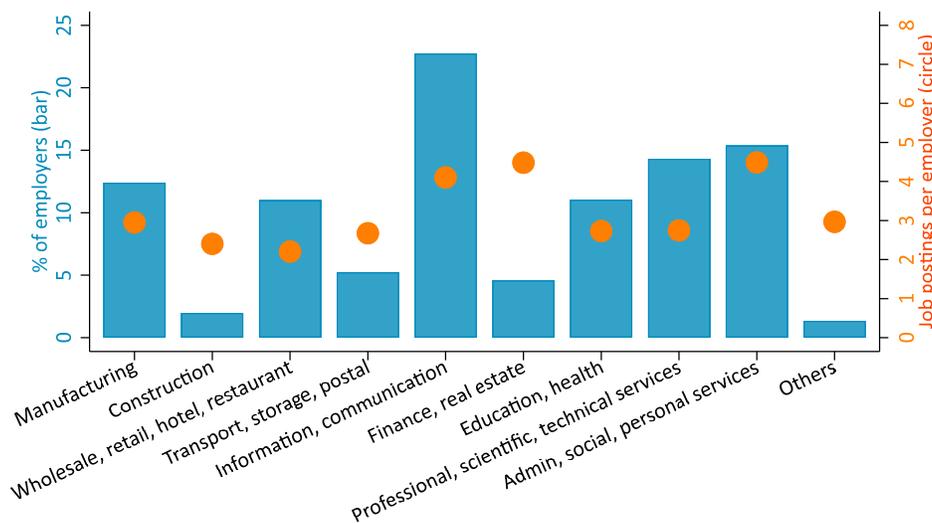


Source: Authors' calculation.

Note: Business's age is truncated at 35.

By industry, information and communication technology (ICT) is the largest industry in the online job portal in terms of the number of employers and job postings (Figure 6). ICT constitutes 23% of the employers, and each employer has posted four jobs on average. The second largest industry is administrative, social, and personal services (15% and 4 jobs), followed by professional, scientific, and technical services (14% and 3 jobs), manufacturing (12% and 3 jobs), and wholesale, retail, hotel, and restaurant (11% and 2 jobs).

Figure 6. Distribution of firms and the number of job postings per firm over industries



Source: Authors' calculation using the data from Rozee.pk.

In sum, the comparison of the characteristics of Rozee.pk data and the labor market data clearly shows that Rozee.pk data represent a high-pay labor market domain where better qualified jobseekers and more productive employers participate. The labor force participants in Rozee.pk are younger and more educated and earn more than the average labor force in Pakistan. Females

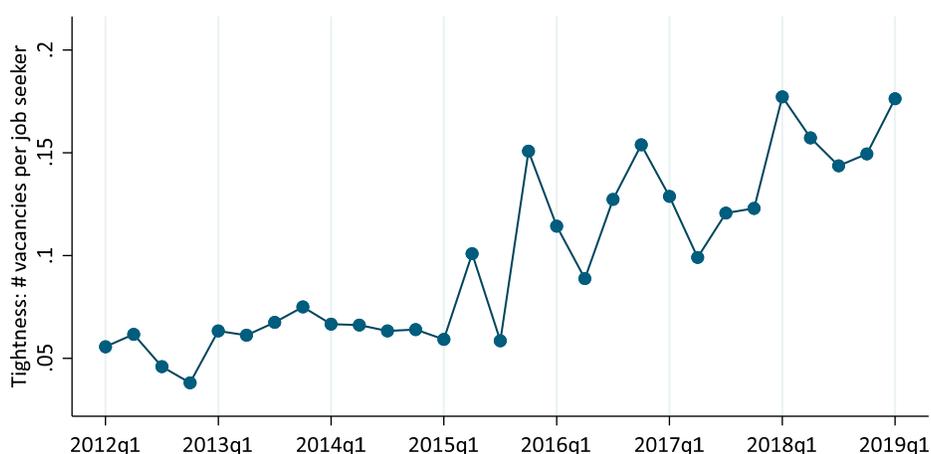
are less likely to participate in the online job portals than males.¹⁴ On the demand side, the employers who use the online job portal are smaller and younger and advertise higher salary jobs than the average employer in Pakistan, and many of them are in ICT.

3. Labor supply and demand

3.1 Labor market tightness and potential mismatches by basic qualifications

In this section, we use a Question and Answer format to navigate key research questions and findings. By exploiting the data feature that the information of both demand and supply sides is available, we measure and look into labor market tightness, which is the number of job vacancies per jobseeker. It shows the relative size of labor demand to labor supply. If the tightness is high, labor demand is strong relative to labor supply, and jobseekers may easily find jobs while employers may have difficulty in finding workers. That is, the market is a seller's market. If the tightness is low, finding jobs is hard for jobseekers, but employers find workers easily. As illustrated in Figure 7, the tightness in Rozee.pk has been increasing, although it is below 0.2 in 2019Q1.¹⁵

Figure 7. Tightness of the online labor market across 2012–2015



Source: Authors' calculation using the data from Rozee.pk.

Note: Shown is the ratio of the number of new vacancies and the number of new jobseekers registered in a given quarter.

Q. What level of education is the most demanded in Rozee.pk, and how easy is it for jobseekers to find jobs at their educational level?

A. Undergraduate and postgraduate education is most demanded in Rozee.pk. However, due to a large number of jobseekers with undergraduate and graduate degrees, the job matching still

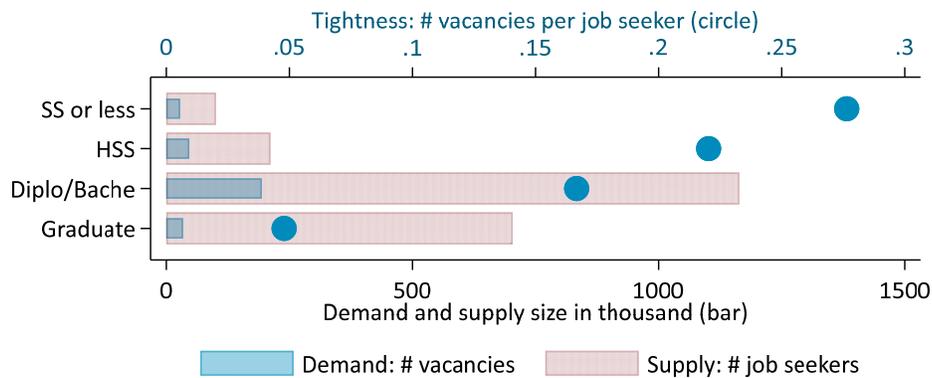
¹⁴ The females who do participate are more educated than the males in the portal. The proportion of those who have postsecondary degrees is 93% and 84% among female and male jobseekers, respectively.

¹⁵ The analysis here does not consider job postings and job seekers outside Rozee.pk; therefore, it does not portray the market tightness of the entire labor market.

remains competitive, and many overqualified jobseekers compete on jobs that require education that is lower than their qualifications.

The number of vacancies by education requirements, illustrated in Figure 8, shows that the online job portal at Rozee.pk is primarily for jobs that require undergraduate and postgraduate education. Of the vacancies, 75% require an undergraduate or postgraduate education. Thus, the portal is mostly for undergraduate and graduate-level jobs and may have a potential for improving job matching of postsecondary education graduates, whose unemployment rate is high (Bossavie et al., 2018). On the other hand, the tightness (i.e., vacancies per jobseeker) by education indicates that finding jobs for graduate degree holders is not easy because the tightness is low at 0.05.¹⁶ (That is, 100 jobseekers with graduate degrees compete for five jobs that require graduate degrees.) By the same token, while the total number of job vacancies that require undergraduate degrees is the greatest, the tightness at the undergraduate level is as low as 0.17 because of the large number of jobseekers who have undergraduate degrees. By contrast, the tightness at the secondary or less education level is the highest at 0.27 since jobseekers whose highest education level is secondary or lower are few.

Figure 8. Labor demand, supply, and market tightness by education



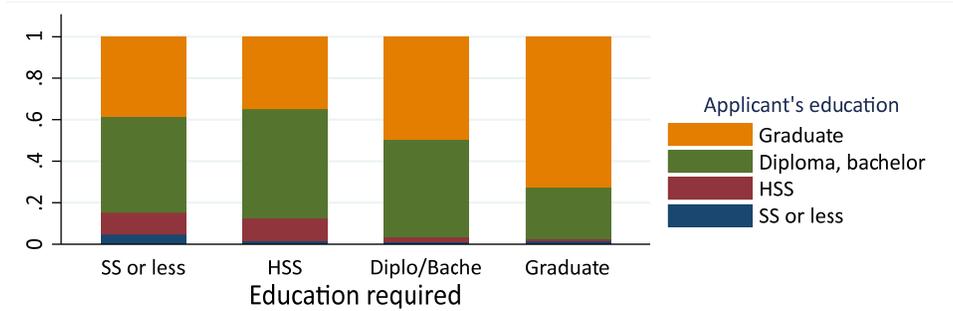
Source: Authors' calculation using the data from Rozee.pk.

Note: The number of vacancies is the total posted since January 2012. The number of jobseekers is the total registered since January 2012.

The lower tightness at the higher education levels may induce high-educated jobseekers to look for low-education jobs, since high-education jobs are less available. In fact, the data show that for jobs that require secondary education or lower level of education, 85% of the applicants have undergraduate and postgraduate degrees (Figure 9). Similarly, for jobs that require a higher secondary education level, 87% of the applicants have undergraduate and postgraduate degrees. Thus, the majority of applicants for secondary-level jobs are overqualified in terms of education. A considerable number of mismatches in terms of education may occur. The key finding from this analysis is that there is an insufficient number of jobs in the online job portal that are suitable for bachelor and graduate degree holders.

¹⁶ In March 2018, the same ratio for the United States was 1.01. The lowest during the last 10 years was 0.15 in July 2009 (<https://fred.stlouisfed.org/series/DHIDFHVTUR>).

Figure 9. Types of applicants by education



Source: Authors' calculation using the data from Rozee.pk.

Note: Shown are the proportions of applicants who have different levels of education.

Q. In which industries and occupations are skills more demanded and supplied, and how is industrial and technological specialization at school relevant to jobs?

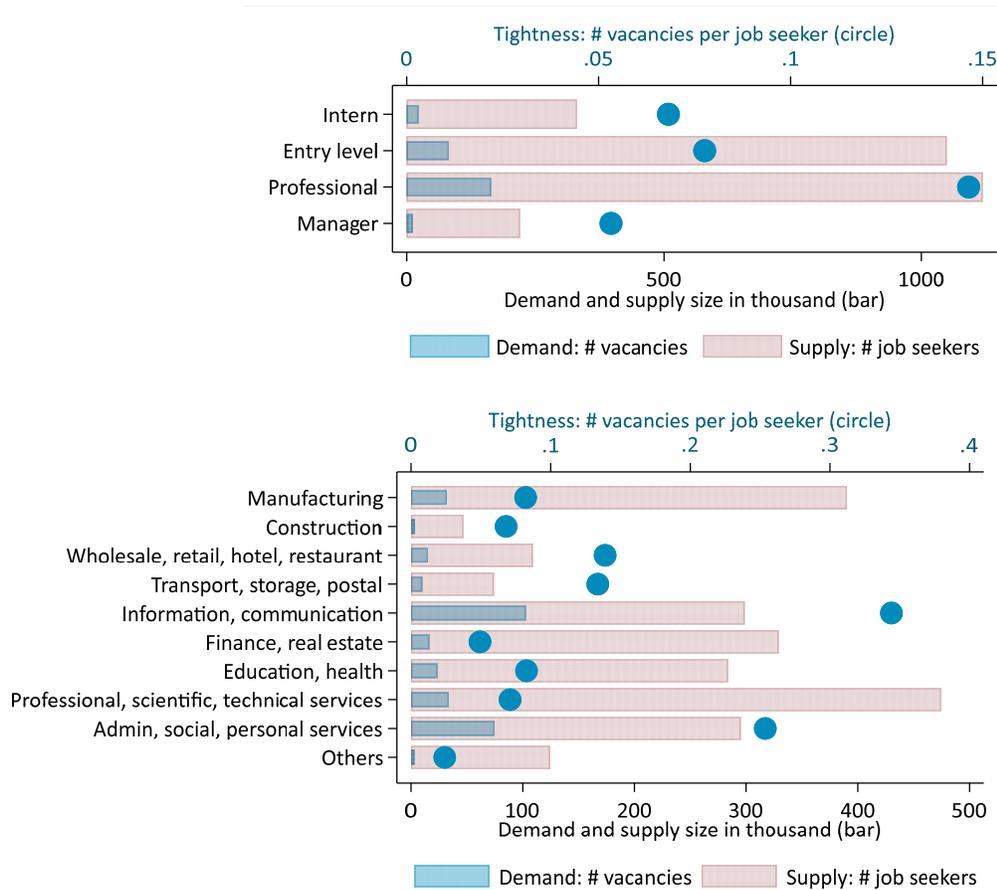
A. In Rozee.pk, the majority of the jobs are the professional level. ICT sector has the largest number of vacancies, because of which the tightness of ICT jobs is the highest, although there are many jobseekers who specialize in ICT.

The labor supply, demand, and market tightness by occupation levels are presented in Figure 10. Occupational levels are grouped into four categories: intern, entry level, professional, and manager. The most prevalent job postings are professional jobs, followed by entry-level jobs. On the supply side too, professional-level and entry-level jobseekers are the majority.¹⁷ However, entry-level jobs are more competitive than professional-level jobs because of limited availability of job vacancies relative to the size of jobseekers, indicating that fresh graduates face more difficulties of finding jobs than those who are already working.

By industry, jobseekers whose self-claimed specialization is either manufacturing, or professional, scientific, and technical services are the two largest groups, whereas job postings in both these industrial categories are relatively few (Figure 10). As a result, the tightness in these two industries is low. On the other hand, ICT sector has the largest number of job vacancies. Although there are many jobseekers in this sector, the tightness for ICT jobs is the highest among all sectors. Thus, jobseekers specialized in ICT may find jobs relatively easily.

¹⁷ Job seekers' occupation levels are self-reported.

Figure 10. Labor demand, supply, and market tightness by occupation levels and industries



Source: Authors' calculation using the data from Rozee.pk.

Note: The number of vacancies is the total number of vacancies posted since January 2012. The number of jobseekers is the total number of jobseekers registered since January 2012.

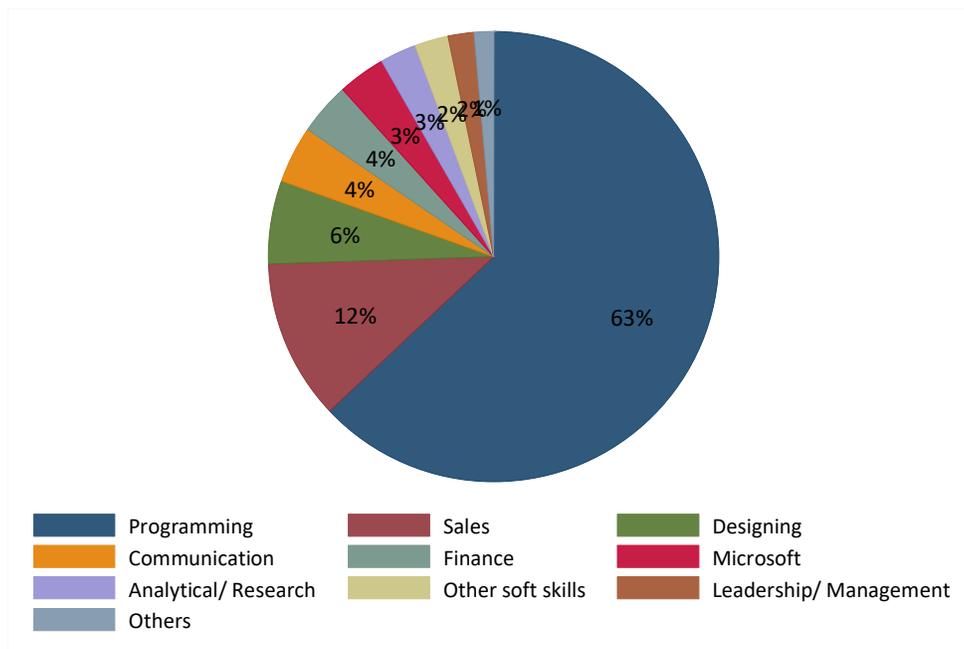
Jobseekers apply for jobs that are outside their industrial specialization. Figure 11 shows the proportion of good-match applicants whose self-claimed specialization matches the industries of jobs they applied for. The proportion of good-match applicants is low in all industries. Of ICT job applicants, 39% are specialized in ICT, and 25% of manufacturing job applicants are specialized in manufacturing. The proportion of good-match applicants is particularly low in the industries for which markets are thin, such as transport and storage (6%), construction (7%), and wholesale, retail, hotel, and restaurant (7%).

Q. What skills are specified in job descriptions? Do these skill demands vary between industry, occupation, or gender of jobseekers?

A. Programming skills are most frequently required (63% of job postings), followed by sales skills (12%). There are differences in skill requirements between occupations and genders. For example, finance skills are most frequently demanded for managers, while communication skills are more frequently demanded for intern- and entry-level positions. By gender, communication and Microsoft skills are more frequently required for female workers.

The 10 most frequently required skill categories are programming, sales, designing, communication, finance, Microsoft, analytics and research, other soft skills,¹⁸ leadership and management and others (Figure 13). Programming skills are required by 63% of all job postings, followed by sales skills (12%), designing skills (6%), and communication skills (4%).

Figure 13. Required skills

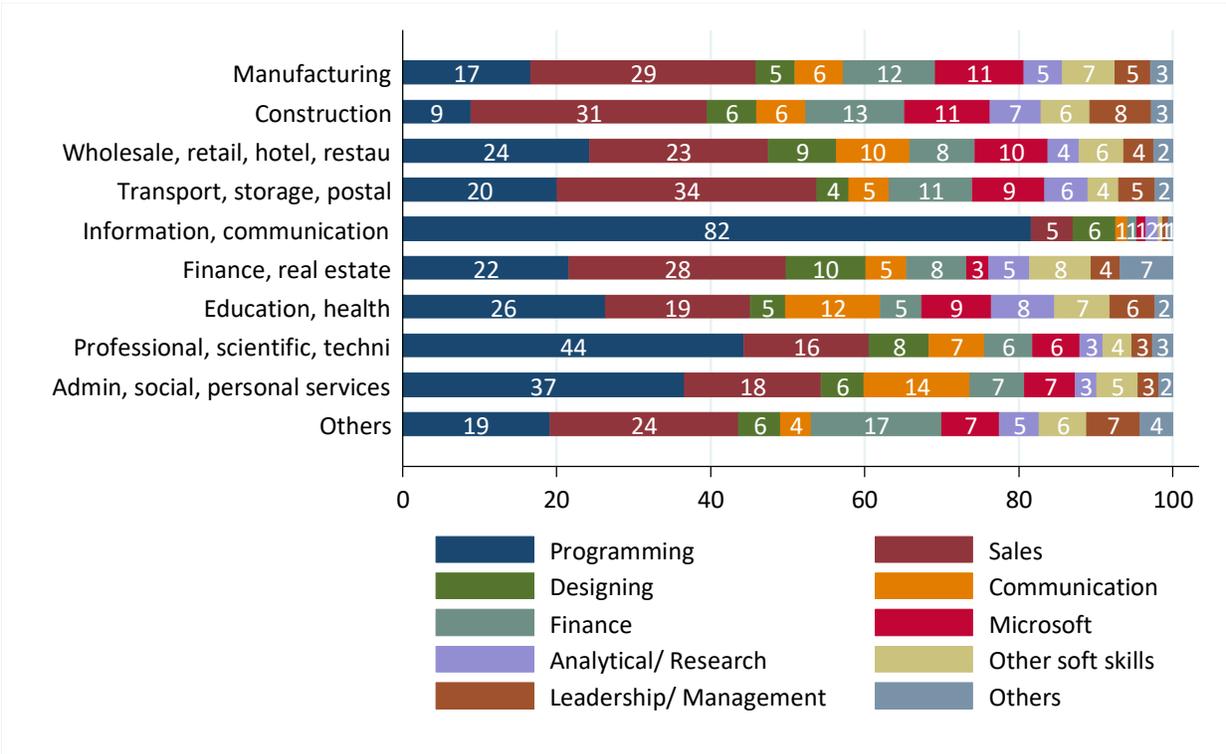


Source: Authors using the data from Rozee.pk.

Figure 14 shows required skills by industries. Programming skills are most frequently required in ICT jobs (82%). Interestingly they are also required in other industries (e.g., 44% in professional, scientific, and technical jobs; 37% in administrative, social, and personal service jobs; and 26% in education and health jobs). This high demand for programming skills in all industries may indicate that employers tend to use online job portals when they look for programming-related skills. Sales’ skills are also frequently required in all industries except ICT.

¹⁸ These include skills such as: time management, team player, multitasking, emergency handling, teamwork.

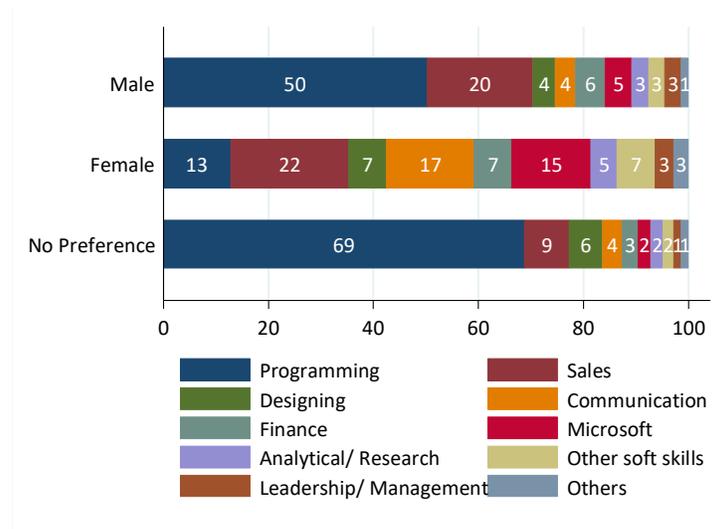
Figure 14. Required skills by industries



Source: Authors using the data from Rozee.pk.

Preferred genders are explicitly indicated in some job postings as discussed in Section 3.4. Figure 15 compares required skills between jobs that prefer different genders. Programming skills are the most frequently required in male-preferred jobs (50%) and gender-neutral jobs (69%), but programming skills are much less frequently required in female-preferred jobs (13%). Compared to male-preferred and gender-neutral jobs, female-preferred jobs are more likely to require sales, communication, Microsoft, and other soft skills.

Figure 15. Required skills by preferred genders



Source: Authors using the data from Rozee.pk.

By applying a machine learning technique to rich text information, such as job titles and descriptions, we categorize each job post and jobseeker by the International Standard Classification Occupations 08 (ISCO08). In our text processing that applies a continuous bag of words (CBOW) model, the job title of a job post is mapped into the most similar occupation title in the ISCO08. Jobseekers’ occupations are identified by mapping their previous job titles into the most similar occupation titles. We can identify the occupations of jobseekers only if they report their previous jobs. Such jobseekers are about half of all jobseekers. The distribution of job posts and jobseekers at the ISCO08 one-digit level is shown in Table 2. The three most frequently demanded and supplied occupations are managers, professionals, and technicians and associate professionals. At the ISCO08 two-digit level (Table A1.1), highly demanded occupations are administrative and commercial managers (ISCO08 = 12), business and administration associate professionals (33), information and communications technology professionals (25), and sales workers (52).

Table 2. Number of job posts and jobseekers by occupations (ISCO08 1 digit)

ISCO-08	Occupation titles	Demand	Supply
1	Managers	312,768	230,938
2	Professionals	439,365	309,331
3	Technicians and associate professionals	324,564	303,325
4	Clerical support workers	46,362	69,733
5	Service and sales workers	118,887	75,340
6	Skilled agricultural, forestry and fishery workers	2,043	4,865
7	Craft and related trades workers	19,334	30,456
8	Plant and machine operators, and assemblers	16,955	25,914
9	Elementary occupations	21,588	15,447

Source: Authors.

Box 1. Occupation and skill categorizations using machine learning

By processing text information in job posts and jobseekers' CVs, we identify their occupations and associated skill categories. Our text processing method basically follows the one developed by Atalay et al. (2018). By estimating a continuous bag of words (CBOW) model, which uses machine learning, job titles and descriptions in job posts are mapped to, respectively, ISCO08 occupations and skill categories. The main output of a CBOW model is to find synonyms for input words and phrases. That is, once a model is estimated, it tells how similar two words are to each other. The model judges the similarity between two words based upon whether these two words are used in "similar" contexts. For example, the model concludes that *realtor* and *real estate agent* are synonyms of one another if they tend to be used near similar words, like *sales*, *purchase*, *buildings*, and *land*.

In our text processing, each job title in the online job data is mapped to the ICSO08 occupation title that our CBOW model finds is the most similar to the job title. A jobseeker's ISCO08 occupation is identified by finding the ISCO08 occupation title that is most similar to the jobseeker's previous job title. Skill requirements in job posts are identified in a similar manner. We first adopt a set of keywords for each skill category from previous studies, namely Deming and Kahn (2018) and Spitz-Oener (2006), then modify these keywords in the context of the Pakistan online job portal by applying a CBOW model, and lastly look at whether a job description includes any of these keywords.

We also identify required skill categories by applying a machine learning method that captures skill-related words in job descriptions. Table 3 presents what percentage of jobs require each skill category. Nonroutine analytic skills are much more frequently required than routine cognitive skills. Nonroutine interactive skills and soft skills—namely social, character, and customer service skills—are frequently required in all occupations. Computer skills are required in about one-third of the jobs of professionals and clerical support workers. Table A1.2 presents, for each skill category, the most skill-intensive occupations at the ISCO08 two-digit level. The occupation of science and engineering professionals is the most intensive occupation in nonroutine analytical skills; teaching professionals in nonroutine interactive skills; assemblers in routine cognitive skills; legal, social and cultural professionals in writing skills; and hospitality, retail and other services managers in people management.

Table 3. Required skill categories by occupations

	Managers	Professionals	Technicians and associate professionals	Clerical support workers	Service and sales workers	Skilled agricultural, forestry and fishery workers	Craft and related trades workers	Plant and machine operators, and assemblers	Elementary occupations
Nonroutine Analytic	0.51	0.56	0.42	0.37	0.35	0.54	0.45	0.42	0.50
Nonroutine Interactive	0.76	0.63	0.66	0.53	0.66	0.75	0.60	0.54	0.67
Nonroutine Manual	0.15	0.14	0.15	0.12	0.11	0.12	0.16	0.18	0.09
Routine Cognitive	0.14	0.12	0.11	0.11	0.06	0.19	0.09	0.11	0.07
Routine Manual	0.14	0.13	0.14	0.12	0.09	0.12	0.15	0.17	0.08
Cognitive	0.28	0.31	0.24	0.42	0.17	0.36	0.20	0.21	0.16
Social	0.45	0.44	0.48	0.34	0.43	0.35	0.36	0.30	0.43
Character	0.54	0.58	0.54	0.47	0.52	0.49	0.48	0.45	0.61
Writing	0.13	0.20	0.12	0.12	0.08	0.09	0.09	0.10	0.09
Customer Service	0.64	0.47	0.60	0.40	0.58	0.47	0.45	0.42	0.53
Project Management	0.39	0.39	0.33	0.32	0.28	0.46	0.35	0.30	0.44
People Management	0.44	0.41	0.40	0.38	0.38	0.56	0.37	0.37	0.52
Finance	0.21	0.15	0.14	0.13	0.08	0.16	0.11	0.13	0.09
Computer	0.23	0.35	0.25	0.33	0.19	0.15	0.20	0.19	0.24

Source: Authors.

Note: Shown are the proportions of job posts that require skills. A job post is considered to require a skill if its job description includes at least one word that is related to a skill.

Box 2. Skill categories

This paper uses two sets of skill categories. The first set consists of nonroutine analytic, nonroutine interactive, nonroutine manual, routine cognitive, and routine manual. These categories are originally put forth by Autor, Levy, and Murnane (2003), who study the impact of computerization on job skill demands in the United States. Their categories are used by Spitz-Oener (2006) to examine changing skill requirements within occupations in Germany. To measure tasks performed by workers, she develops a mapping from specific task words to skill categories as shown in Table 4. Our analysis builds on her mapping. Applying the machine learning method described in Box 1, we identify synonyms of original task words in the context of the online job portal, and then look at whether a job description includes any of the original words or their synonyms.

Table 4. Skill categories and related tasks in Spitz-Oener (2006)

Skills	Tasks
Nonroutine analytic	Researching, analyzing, evaluating and planning, making plans/constructions, designing, sketching, working out rules/prescriptions, and using and interpreting rules
Nonroutine interactive	Negotiating, lobbying, coordinating, organizing, teaching or training, selling, buying, advising customers, advertising, entertaining or presenting, and employing or managing personnel
Nonroutine manual	Repairing or renovating houses/apartments/machines/vehicles, restoring art/monuments, and serving or accommodating
Routine cognitive	Calculating, bookkeeping, correcting texts/data, and measuring length/weight/temperature
Routine manual	Operating or controlling machines and equipping machines

Source: Adapted from Spitz-Oener (2006).

The second set, consisting of nine skills listed in Table 5, is based on Deming and Kahn (2018). The left column in the table presents skill categories, and the right column lists keywords and phrases related to corresponding skill categories.

Table 5. Skill categories based on Deming and Kahn (2018)

Skills	Keywords and phrases
Cognitive	Problem solving, research, analytical, critical thinking, math, statistics
Social	Communication, teamwork, collaboration, negotiation, presentation
Character	Organized, detail oriented, multitasking, time management, meeting deadlines, energetic
Writing	Writing
Customer service	Customer, sales, client, patient
Project management	Project management
People management	Supervisory, leadership, management (not project), mentoring, staff
Financial	Budgeting, accounting, finance, cost
Computer	Computer, spreadsheets, common software (e.g., Microsoft Excel, PowerPoint)

Source: Adapted from Deming and Kahn (2018).

3.3 Salaries

This section examines how salaries are associated with job characteristics and qualifications.

Q. How do salaries differ depending on required qualifications, occupations and industries?

A. Regression analysis finds that higher education degrees, more experience and higher occupational levels are associated with higher salaries. By industry, the ICT sector advertises the highest average salaries. By firm size, larger firms advertise higher salaries.

We conduct a regression analysis of the correlations between salaries and job characteristics. The regression results are summarized in Figure 16. The dots and numbers in the figure indicate differences in salaries from a reference category. For example, in the case of industries, the reference category is manufacturing, and the dots and numbers indicate how much the average salaries in the other industries are different in proportion from the average salary in manufacturing.

Figure 16. How salaries in job postings differ across job types and qualifications



Source. Authors

Note. Shown are regression coefficients and their 95% confidence intervals. The dependent variable is log of the midpoint of the salary range of a posted job. The other independent variables included are dummies for preferred genders, employment type dummies, time shift dummies, the number of vacancies, and year-month dummies.

There are six key findings from this regression analysis.

- *Finding 1: Posted salaries increase with education requirements.* Compared to job postings that require only secondary education or less, those that require higher secondary education, undergraduate degrees, and postgraduate degrees advertise 1%, 17%, and 23% higher salaries, respectively.
- *Finding 2: Salaries increase with increased experience requirements.* Compared to job postings that require zero year of experience, those that require 0.1–2 years, 3–5 years, 6–9 years, and 10 or more years of experience advertise 9%, 46%, 93% and 110% higher salaries, respectively.
- *Finding 3: Salaries and skill requirements are significantly correlated.* Job postings that require cognitive, social, character, customer service, writing, project management, people management, and computer skills advertise, respectively, 1%, 5%, 5%, 7%, 2%, 5%, 2% and 9% higher salaries than those that do not require these skills. Interestingly, financial skills are negatively correlated with salaries. Nonroutine analytic, nonroutine manual, routine manual skills are positively correlated with salaries, while nonroutine interactive and routine cognitive skills are negatively correlated.
- *Finding 4: Salaries increase with occupation levels.* Entry-level jobs, professional jobs and manager job postings advertise 13%, 40% and 67% more than intern jobs, respectively.
- *Finding 5: Salaries are higher at larger firms.* Compared to micro-size firms, medium-, large-, and mega-size firms advertise 16%, 16%, 29% higher salaries.
- *Finding 6: Salaries are the highest in ICT (21% higher than manufacturing jobs) and lowest in the education and health sector (9% lower).*

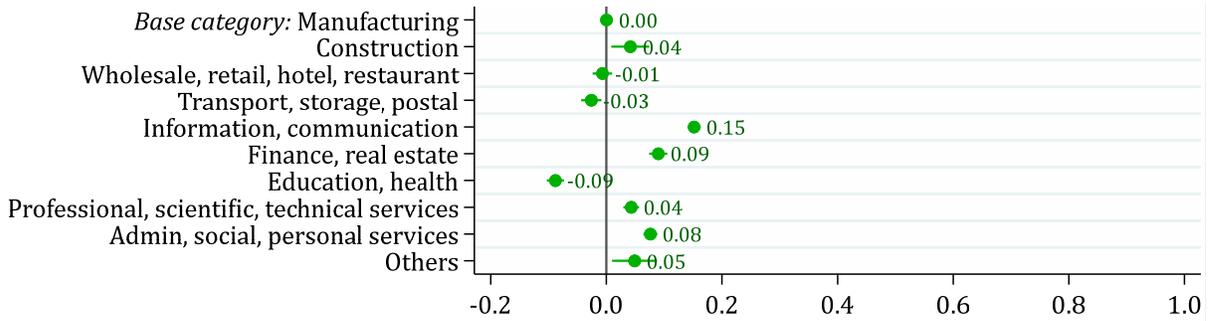
Q. For fresh college graduates, which industry advertises high initial salaries and may provide high salary growth?

A. ICT and administrative, social, and personal services advertise high salaries. In these sectors, the salary growth profile looks high too.

Given that approximately two million people enter the labor market annually, it is important to understand job opportunities for those fresh labor market entrants.¹⁹ Figure 17 shows differences between industries in salaries for fresh college graduates. The ICT sector advertises the highest salaries in their job postings (15% higher than manufacturing); finance and real estate the second (9% higher than manufacturing); and administrative, social, and personal services the third (8% higher than manufacturing). Advertised salaries are the lowest in the education and health sector (9% lower than manufacturing).

¹⁹ Note that the postsecondary degrees analyzed here include both professional and academic degrees.

Figure 17. Which industries offer higher salaries to college graduates?

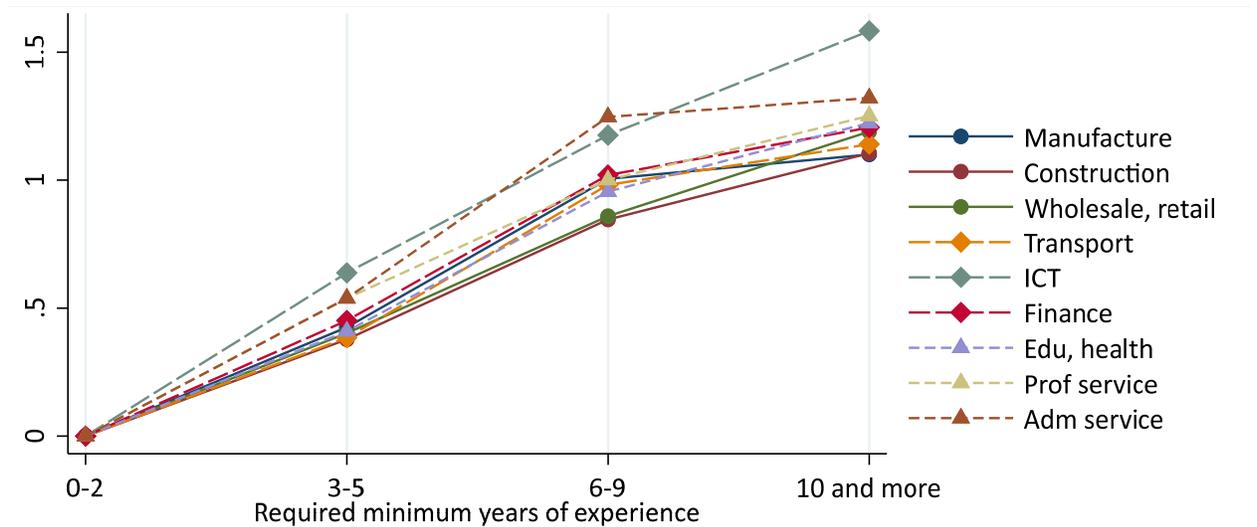


Source: Authors.

Note: Shown are regression coefficients and their 95% confidence intervals. The sample used is the job postings that require undergraduate degrees and 0–2 years of experience. The same regression specifications reported in **Error! Reference source not found.** are applied, that the education dummies and experience dummies are not controlled for.

Figure 18 looks at the return to experience. The ICT sector (which advertises the highest entry-level salaries, as shown in Figure 17) has the highest salary growth trajectory, so the ICT sector may be a good career choice, in terms of salaries, for postsecondary school graduates. By contrast, the education and health sector (where the initial salaries are the lowest) has a relatively low salary growth trajectory.

Figure 18. Returns to experience by industries



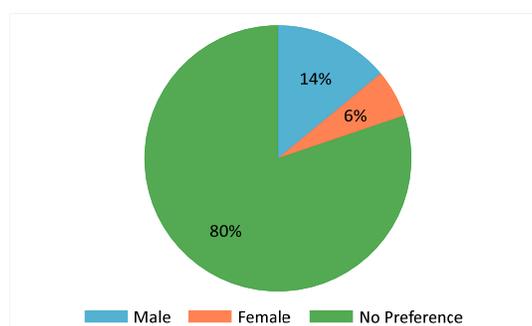
Source: Authors.

Note: Shown are salary differences, within a given industry, between job postings that require different levels of professional experience. These differences are estimated by regressions separately run for a subsample of job postings in each industry that require an undergraduate education. The controls are dummies for preferred genders, firm size dummies, employment type dummies, time shift dummies, the number of vacancies, and year-month dummies.

3.4 Gender preference in job postings

Hiring practices based on gender are not uncommon in Pakistan, and gender preferences of employers can be explicitly signaled in job postings in Rozee.pk. Every job posting has the option to indicate gender preference, i.e., male or female. While 80% of jobs do not indicate any gender preference, 14% and 6% of job postings indicate that they prefer, respectively, men and women (Figure 19). This indication of gender preference is effective in attracting preferred-gender applicants, as shown in Figure A1.3. The proportion of male applicants is 97% for male-preferred job vacancies, while the proportion of female applicants is 51% for female-preferred vacancies.

Figure 19. Percentage of gender-earmarked vacancies



Source: Authors' calculation using the data from Rozee.pk.

Q. What type of jobs prefer male and female workers? Are there salary differences associated with gender preference?

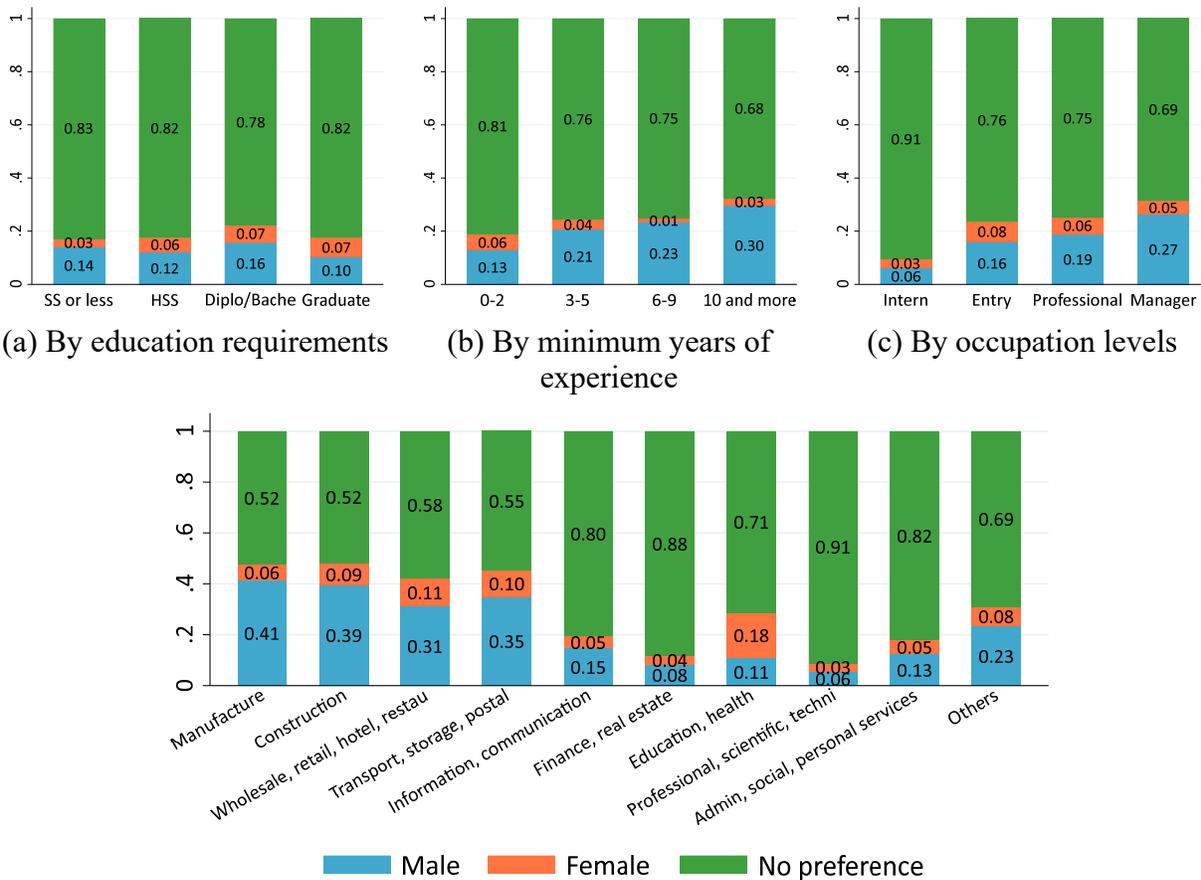
A. Jobs that require longer professional experience and are at higher occupation levels are more likely to prefer males. Job postings that indicate gender preference, irrespective of which gender is preferred, advertise lower salaries than the job postings that do not indicate a preference.

We examine what types of job postings indicate gender preferences. We also examine how gender preferences are associated with advertised salaries (Figure 20). Key findings are summarized below.

- *Finding 1: Female workers are more likely to be preferred in jobs requiring higher education. 7% of the jobs requiring bachelor and graduate-level degrees prefer female workers, whereas only 3% of the jobs requiring the secondary and less education level prefer females.*
- *Finding 2: Jobs that require longer years of experience and managerial jobs tend to prefer men. 30% of the jobs that require 10 or more years of experience prefer males, whereas 13% of the jobs requiring zero to two years of experience do so. 27% of managerial jobs prefer men, whereas 6% of intern jobs prefer men.*

- Finding 3: Professional, scientific, and technical sector, and finance sector are the most gender-free.* In these two sectors, about 90% of jobs do not have gender preferences. On the other hand, 31 to 41% of the jobs in manufacturing; construction; wholesale, retail, hotel, and restaurant; and transport, storage, and postal sectors prefer male workers. Female workers are the most frequently preferred in education and health sector.
- Finding 4: Regarding ISCO08 occupations, ICT jobs are less common among female-preferred jobs than male-preferred ones. Teaching professions are more common among female-preferred jobs than male-preferred jobs.* We look into occupational compositions at the ISCO08 two-digit level in each of the gender-free, male-preferred, and female-preferred job categories (Table 6). Among both male- and female-preferred jobs, administrative and commercial managers (ISCO08 = 12) and business and administration associate professionals (ISCO08 = 33) are the two most common. Information and communications technology professionals (ISCO08 = 25) is the third most common among male-preferred jobs, constituting 15%, but is less common among female-preferred jobs, constituting 9%. The occupation of teaching professionals (ISCO08 = 23) is relatively common among female-preferred jobs (7%), while it constitutes only 2% of the male-preferred jobs.

Figure 20. What type of jobs prefers males or females?

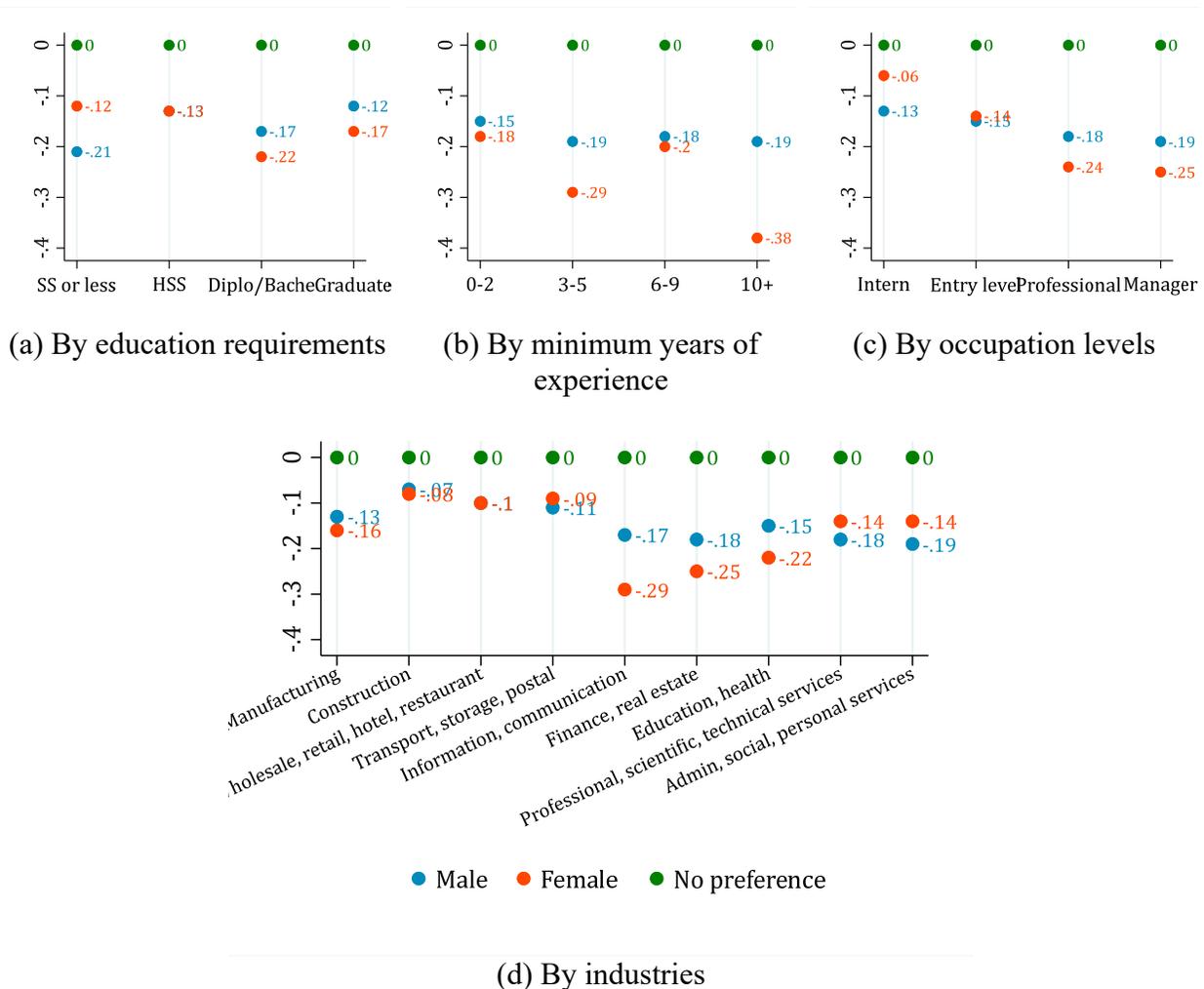


(d) By industries

Source: Authors' calculation using the data from Rozee.pk.

- Finding 5: Job postings with gender preference offer, on average, 15% lower salaries than those without any gender preference, after controlling for other job characteristics.* When male- and female-preferred jobs are compared, male-preferred jobs pay slightly higher than female-preferred jobs except for a few cases, but they universally pay lower than the jobs having no gender preference. These findings are overall consistent across different industries and education and experience requirements (Figure 21). Chowdhury et al. (2018) find a similar pattern in an Indian online job portal, where gender-preferred jobs offer 6% lower salaries than jobs with no gender preference, and jobs that prefer women offer 10% lower salaries than jobs that prefer men.

Figure 21. Salaries and gender preference by education and experience requirements, occupation levels, and industries



Source: Authors' calculation using the data from Rozee.pk.

Table 6. Occupations (ISCO08 two-digit) by gender preferences

ISCO	Occupation Titles	No preference	Male	Female
11	Chief executives, senior officials and legislators	0.006	0.007	0.007
12	Administrative and commercial managers	0.196	0.239	0.190
13	Production and specialised services managers	0.025	0.027	0.022
14	Hospitality, retail and other services managers	0.007	0.010	0.007
21	Science and engineering professionals	0.062	0.067	0.049
22	Health professionals	0.008	0.007	0.011
23	Teaching professionals	0.024	0.015	0.066
24	Business and administration professionals	0.053	0.045	0.056
25	Information and communications technology professionals	0.163	0.146	0.092
26	Legal, social and cultural professionals	0.036	0.022	0.035
31	Science and engineering associate professionals	0.010	0.017	0.007
32	Health associate professionals	0.031	0.029	0.044
33	Business and administration associate professionals	0.190	0.153	0.209
34	Legal, social, cultural and related associate professionals	0.012	0.010	0.012
35	Information and communications technicians	0.009	0.010	0.010
41	General and keyboard clerks	0.022	0.022	0.030
42	Customer services clerks	0.008	0.009	0.019
43	Numerical and material recording clerks	0.002	0.003	0.001
44	Other clerical support workers	0.001	0.001	0.001
51	Personal service workers	0.008	0.021	0.010
52	Sales workers	0.069	0.067	0.067
53	Personal care workers	0.012	0.008	0.020
61	Market-oriented skilled agricultural workers	0.001	0.001	0.001
62	Market-oriented skilled forestry, fishery and hunting workers	0.000	0.000	0.000
63	Subsistence farmers, fishers, hunters and gatherers	0.000	0.000	0.000
71	Building and related trades workers, excluding electricians	0.004	0.005	0.004
72	Metal, machinery and related trades workers	0.002	0.004	0.001
73	Handicraft and printing workers	0.001	0.001	0.001
74	Electrical and electronic trades workers	0.003	0.006	0.003
75	Food processing, wood working, garment and other craft and related trades workers	0.005	0.006	0.007
81	Stationary plant and machine operators	0.004	0.009	0.004
82	Assemblers	0.000	0.001	0.000
83	Drivers and mobile plant operators	0.006	0.016	0.005
91	Cleaners and helpers	0.001	0.001	0.001
92	Agricultural, forestry and fishery labourers	0.000	0.000	0.000
93	Labourers in mining, construction, manufacturing and transport	0.001	0.001	0.001
94	Food preparation assistants	0.002	0.004	0.002
95	Street and related sales and service workers	0.000	0.000	0.000
96	Refuse workers and other elementary workers	0.013	0.007	0.007

Source: Authors.

Note: The proportions of occupations in each gender preference are presented.

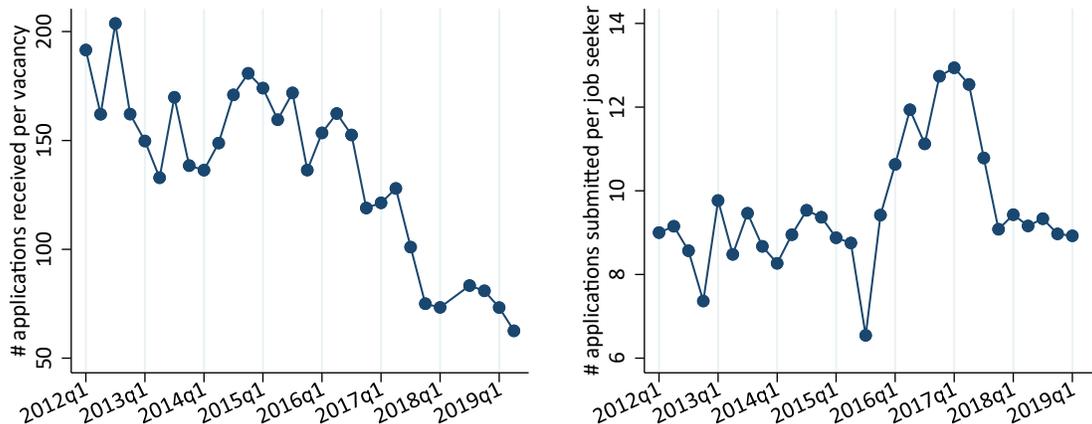
4. Job application and selection

4.1 Applications

This section analyzes job market transactions, namely application and selection. One of the unique data features of online job portals is transaction data, which cannot be obtained from labor force surveys. Over the period from 2012, the number of applications per job posting is

180. An average active jobseeker submitted around 10 applications per quarter (Figure 22). The next set of research questions focuses on the behavioral patterns of jobseekers and employers in application and selection.

Figure 22. Number of applications per vacancy and jobseeker over time



Source: Authors' calculation using the data from Rozee.pk.

Note: The activeness of jobseekers is defined as that they apply for at least one job in a given quarter.

Q. How fast do job postings receive applications? What types of job postings receive more applications?

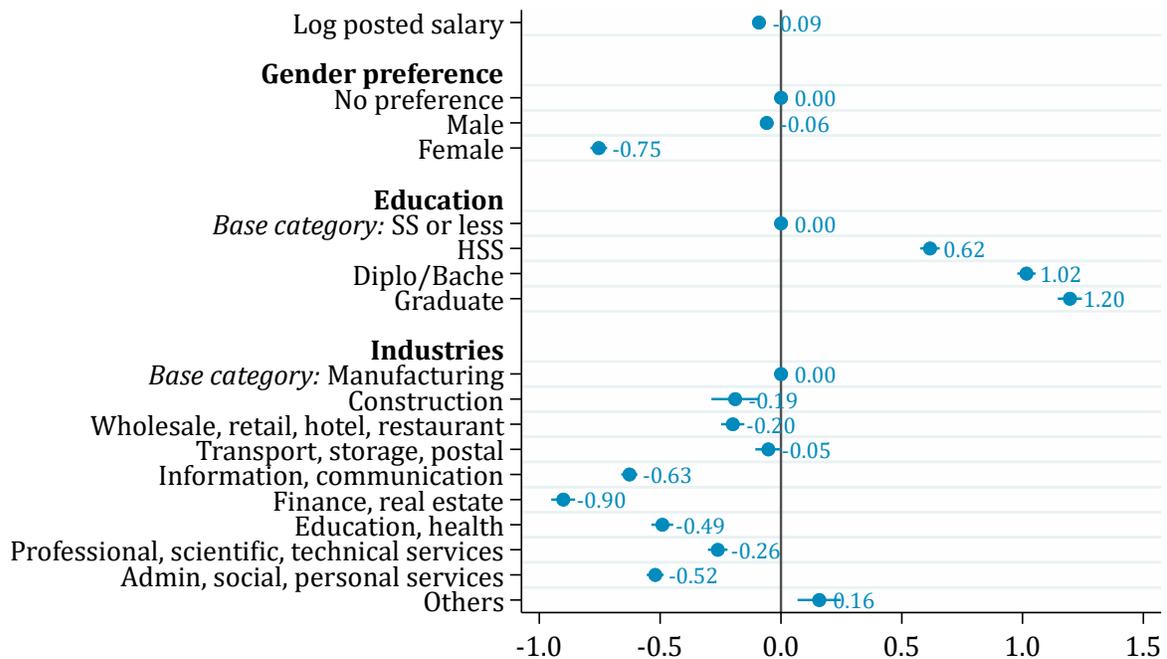
A. The online job market is fast. A job posting receives nearly 100 applications on average in two weeks. Jobs that require a higher level of education, do not indicate gender preference, and are in manufacturing sector receive a greater number of applications. Job postings advertising higher salaries receive a smaller number of applications.

- *Finding 1: Many applications were received immediately after jobs were posted. 80% of applications were received in 14 days, and 90% in 25 days. This means that out of 180 applications to an average vacancy, more than 150 applications were received within 25 days. This shows how fast the online job market is.*
- *Finding 2: Job postings that do not indicate gender preference receive more applications. The average number of applications is 6% and 75% fewer for, respectively, male- and female-preferred jobs than gender-free jobs. This result shows the effectiveness of signaling gender preference, particularly preference for female workers.*
- *Finding 3: Jobs that require higher education receive more applications. Jobs requiring undergraduate and postgraduate degrees receive more than double the applications to the jobs requiring secondary or less education.*
- *Finding 4: Manufacturing jobs receive the greatest number of applications. Jobs in finance and real estate and ICT receive the least. The number of applications per job post*

is the highest in manufacturing sector. The number of applications per job post is 90% and 63% fewer in finance and real estate sector and ICT sector, respectively.

- *Finding 5: Higher salaries advertised in job postings are negatively correlated with the number of applications received. Job postings that advertise double salaries receive 9% less applications.²⁰ This negative correlation is interesting. This may be because higher salaries may inform jobseekers that job requirements are higher or because jobseekers may expect higher-salary jobs to be more competitive, so they do not apply for them.*

Figure 23. What type of job postings receives more applications



Source: Authors.

Note: The dependent variable is the log of the number of applications a job post received. The other controls are dummies for experience requirements, occupation levels, company size, employment types, time shift, subscriptions of add-on services and year-months.

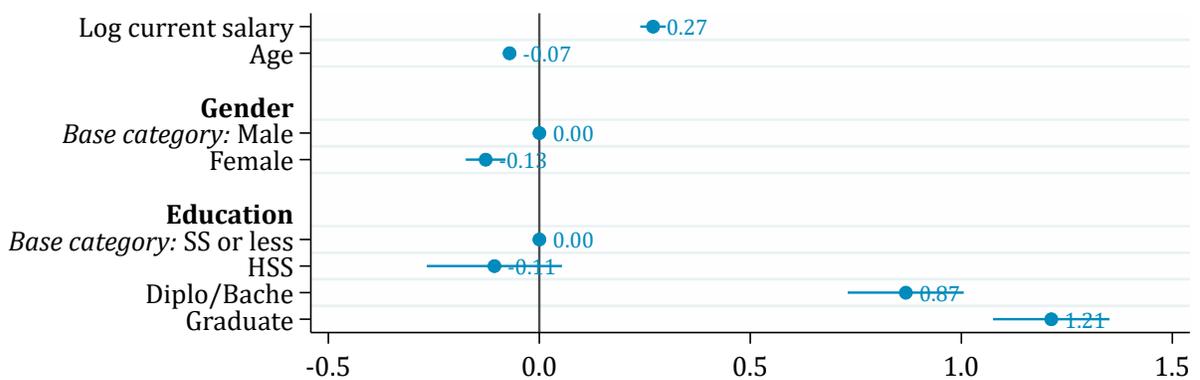
Q. What are the behavioral patterns in jobseekers' activities?

A. Jobseekers who are male, young, and with higher current salaries and educational degrees submit more applications. Many jobseekers apply for jobs that offer less than their desired salary level.

²⁰ The salary range is not always made public to job seekers. Out of the job postings where employers enter salary information in Rozee.pk's website, approximately 50% keep the salary information confidential from job seekers. The regression result, however, is similar whether salary information is disclosed or not is taken into account.

To understand jobseekers' behavior, it is useful to analyze how many applications they submit to what kinds of job postings. The regression reported in Figure 24 examines who submits more applications. It finds that female jobseekers apply to 13% fewer jobs than male jobseekers. This may be because females have a better chance of being selected with fewer applications (which is discussed in a later analysis) or they cannot afford to apply for as many jobs due to tighter resource constraints. One-year younger jobseekers submit 7% more applications. Jobseekers with undergraduate and postgraduate degrees submit approximately twice as many applications as jobseekers with secondary or less education. As seen earlier, labor market tightness is low for graduates with higher education degrees, so highly educated jobseekers may need to apply for more jobs. It may also be the case that highly educated jobseekers have more job openings they qualify for. Lastly, jobseekers who currently earn higher salaries submit more applications. (Jobseekers with double salaries submit 18% more applications.)

Figure 24. Who submits more applications



Source: Authors.

Note: The other controls are experience dummies, industries, marital status, and year-month dummies.

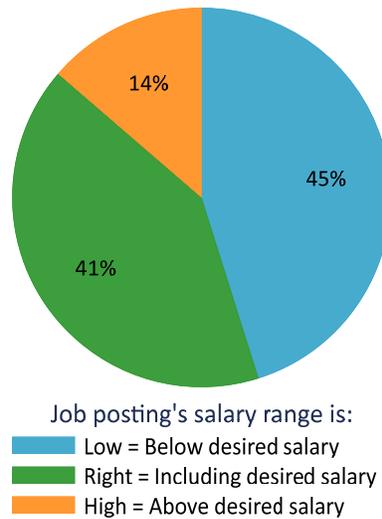
We compare jobseekers' desired salaries and the salaries advertised in the job postings they apply for (Figure 25). Jobseekers express their desired salaries in their profiles on Rozee.pk. Jobseekers apply for the jobs that pay lower than their desired salaries in 45% of the cases. The difference between the advertised and desired salaries is often as large as 40%.²¹ This can be accounted for by the fact, demonstrated in an earlier section, that highly educated jobseekers apply for jobs requiring lower education. Another possible reason why jobseekers often apply for lower paying jobs is their risk aversiveness.²² It is also worth noting that in 14% of cases, workers apply for jobs that pay higher than their desired salaries. This shows that they still try getting higher paying jobs while they mainly target not-so-high paying jobs.²³

²¹ In half of the cases where workers apply for jobs that pay less than their desired salaries, the maximum salaries posted are more than 40% lower than their desired salaries.

²² It may also be the case that self-declared desired salaries are higher than what they truly desire.

²³ (a) We have a similar finding when we compare job postings' salary ranges and applicants' current salaries. (b) These statistics are not substantially different between applicants' gender.

Figure 25. Job posting's salaries relative to applicants' desired salaries



Source: Authors' calculation using Rozec.pk data.

4.2 Selection: Short listing

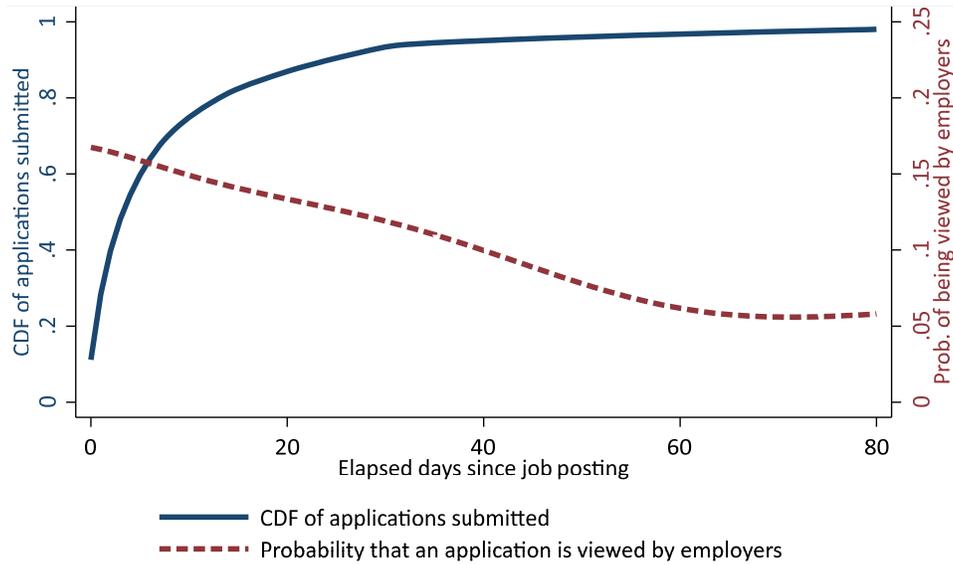
This section analyzes how employers select workers. This analysis is rare and nearly impossible with traditional labor force survey data. An exceptional example is Nomura et al. (2017), who examine what affects the probability of being shortlisted by using online job portal data in India.

Q. What type of applications are more likely to draw employers' attention and be shortlisted?

A. Applicants whose qualifications match job postings have a higher probability of being shortlisted. Both underqualified and overqualified candidates have a lower probability of being shortlisted.

Figure 26 illustrates the timing of applications submitted (blue solid line) and the probability of an application being viewed by employers (red dash line). As already discussed, 80% of applications were received in 14 days, and 90% were received in 25 days. The probability of being viewed decreases as an application is submitted later. If an application is submitted on the same day of job posting, the probability of being viewed is 17%, but the probability is 13% and 6% if submitted 20 and 60 days later, respectively. The timing of application matters in short listing.

Figure 26. Application timing and results



Source: Authors.

Note: The data used are a subsample of the transactions associated with jobs that were posted in or before 2017Q1. The red line (the probability of an application being viewed by employers) is based on an estimate of a seven-order polynomial function of elapsed days.

Employers' decisions on short listing are examined by regressions that control for job posting fixed effects (Figure 27). Since the regressions control for job posting fixed effects, the analysis compares the probability of being shortlisted between applications submitted to the same jobs.²⁴ The key findings are as follows:

- *Finding 1: In the case of job postings that do not indicate gender preference, female applicants are 3.3 ppt more likely to be shortlisted than males.* This difference is economically significant given that the mean probability of being shortlisted is 9.5%.
- *Finding 2: In the case of job postings that indicate gender preference, matched genders lead to a higher probability of being shortlisted.* In the case of male-preferred jobs, male applicants are 1.4 ppt more likely to be shortlisted than female applicants. In the case of female-preferred jobs, female applicants are 11 ppt more likely to be shortlisted than male applicants. It may be surprising that gender-unmatched applicants are shortlisted in some cases.
- *Finding 3: Later applications are less likely to be shortlisted.* If elapsed days from a job posting to the timing of application is twice as long, the probability decreases by 1.1 ppt.

²⁴ Information as to whether an application was shortlisted was voluntarily updated by employers. In fact, many applications seem to be not updated. Thus, we apply the following data restrictions. First, we use only the job postings that were posted in or before 2018Q1. Those job postings must have completed short listing by April 2019. Then, we restrict the sample to the job postings for which at least one application was shortlisted. In other words, if none of the applications to a certain job posting are recorded as having been shortlisted, we exclude the job posting and all the applications to the job posting. The above two steps make the subsample for the analysis here. Regarding application status, we interpret that an application was not shortlisted if the status of the application was not updated, and another application was shortlisted.

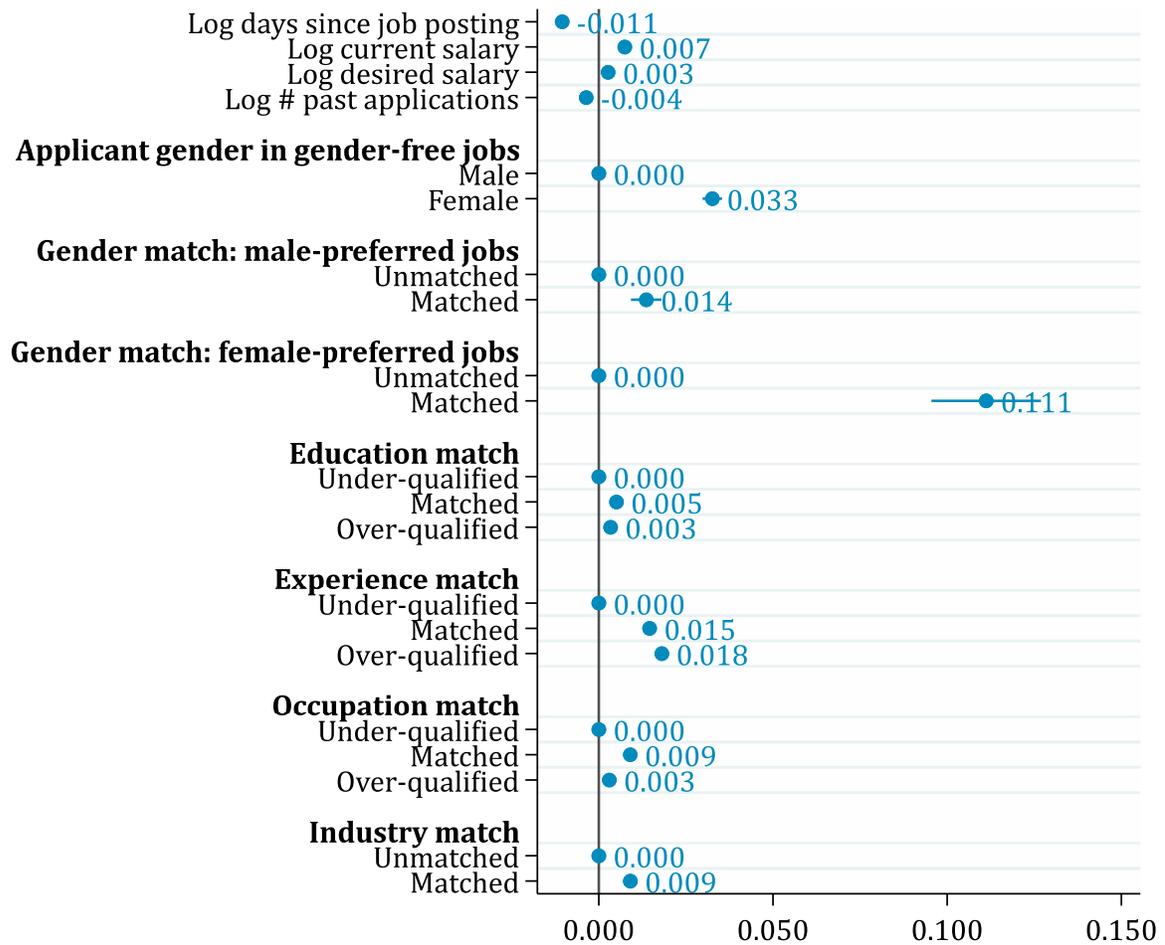
- *Finding 4: Good matches in terms of education increase the probability of being shortlisted.* Compared to the case where applicants' education levels are lower than education requirements, the probability of being shortlisted is 0.5 ppt higher if applicants' education exactly matches the required level. Compared to these just-qualified applicants, overqualified applicants are no more likely to be shortlisted.
- *Finding 5: Good matches in terms of years of experience increase the probability of being shortlisted.* Compared to underqualified applicants in terms of years of experience, those whose years of experience exactly match the minimum years of experience required are 1.5 ppt more likely to be shortlisted. The shortlisted probability is not statistically significantly different between just-qualified and overqualified applicants.
- *Finding 6: Occupational matches affect short listing.*²⁵ Compared to underqualified applicants, whose self-reported occupation levels are below jobs' levels, the probability of being shortlisted is 0.9 ppt higher for just-qualified applicants. Interestingly, the probability for overqualified applicants is about the same as that for underqualified applicants and smaller than that for just-qualified applicants.
- *Finding 7: Matches in terms of industries are also valued in employers' short-listing decisions.* If applicants' self-claimed industrial specialization matches jobs' industries, the applicants are 0.9 ppt more likely to be shortlisted than otherwise. This effect of industrial good matches is the strongest in wholesale, retail, hotel, and restaurant sector (2.5 ppt), and small in finance and real estate services (0.3 ppt) and professional, scientific, and technical services (0.5 ppt).²⁶

In sum, the results show that applicants who match requirements by job postings are more likely to be shortlisted. Not only does underqualification lead to a lower probability of being shortlisted, but overqualification does so.

²⁵ In the regression, occupation levels are categorized into either: (a) intern and entry level or (b) professional and manager.

²⁶ See Figure A1.4.

Figure 27. Who is shortlisted?



Source: Authors.

Note: Shown are regression results that examine how much characteristics of applicants/applications affect the probability of being shortlisted or interviewed. The regression controls for job postings fixed effects. The dependent variable is a dummy indicating that an application is shortlisted or interviewed. The other controls are marital status dummies and job postings fixed effects. The mean probability of being shortlisted is 9.5%. The standard errors are clustered at the job posting level.

5. Conclusions

This paper provides unique, granular pictures of labor markets by using novel online job data. The data of Rozee.pk enable us to analyze labor markets in real time on both employer and worker sides and understand skill demands and supply based on rich text information.

The online job portal examined in this paper represents a high-skill segment of the labor market in Pakistan. Job postings in the job portal offer higher salaries than the national average, and jobseekers are younger and better educated than the average labor force in the country. This focus on a high-skills segment is relevant to tackle the ongoing labor market issues of Pakistan such as the youth bulge and the higher unemployment rate of better educated people.

The paper examines various aspects of the labor market, including skills demand and supply, returns to skills, gender preference, and jobseekers' and employers' behavior in job application and selection. The paper presents findings in the Question and Answer format.

One of the key findings is that there is an insufficient number of jobs in which bachelor and graduate degree holders are expected to use their skills obtained from higher education. The labor market tightness at the postsecondary degree level is low. Because a new batch of fresh graduates of colleges and postsecondary education enter the labor market every year, there is more competition for entry-level jobs than for professional-level jobs that may focus on more mid-career experienced workers. Due to the limited availability of job vacancies relative to the number of jobseekers, fresh college graduates tend to face more difficulties in finding jobs than those who are already working. However, this varies by industry. For example, jobseekers with an ICT specialization tend to find it easier to find jobs through Rozee.pk than jobseekers with other industrial specializations because the tightness of the ICT sector is highest among all sectors. This is because the relative number of available jobs to the number of jobseekers is larger than other sectors.

The analysis of short listing shows that matching between applicant qualifications and job requirements is important and that overqualification does not necessarily lead to any advantages. Rozee.pk allows employers to express specific skills requirements for jobs in their job descriptions. A keyword analysis of skills requirements shows that the employers tend to specify programming-related skills (63% of expressed keywords) followed by sales-related skills (12%). The needs for programming skills, or more broadly, ICT skills, are validated by other parameters in this paper. The wage offers in terms of both initial wage and wage growth trajectory show that ICT is one of the most attractive sectors for postsecondary education graduates. Programming-related skills are demanded not only in the ICT sector, but also in all other industries as well. In sum, matching skills in terms of skill levels (or educational qualifications) and industry specialization is one of the important findings of this analysis. On the contrary, many overqualified applicants tend to apply for jobs that require lower educational requirements than their educational backgrounds due to low tightness of the job market for the highly educated segments.

Skills mismatch is a prevalent phenomenon in many countries, and such mismatches can be in forms of educational levels, industry specialization, and quality of skills. In this regard, job matching processes and methods have considerable room for improvement. While the higher skills segment of the job market seems to have opportunities to learn about job availabilities through Rozee.pk, the service does not cater to the middle- to low-skilled population. Recent technological advancement, with the spread of mobile phones or smartphones, and accessibility of middle-income and lower-income households have drastically improved worldwide. In Pakistan mobile phone subscriptions per 100 people increased from 8.3 in 2005 to 73.4 in 2017.²⁷ As seen in the case of India, analyzed by Nomura et al. (2017), the service catered to middle- to lower-skilled population can be a possible opportunity for Pakistan to improve the overall skills matching for the economy. The analysis in this paper indicates that the costs are high for many jobseekers, so they cannot afford sufficient job searches. Female jobseekers may

²⁷ World Bank Open Data, October 18, 2018.

lack resources and thus can only apply for a few jobs. Because of high job search costs, poor workers as well as low-skilled workers may use informal job searches through social networks, which is cheap but likely to lead to poor match quality (Matsuda and Nomura, 2018).

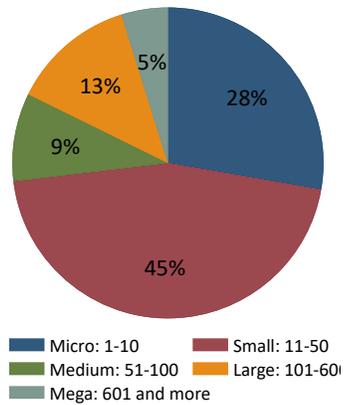
As this paper has demonstrated, job portals can provide very useful knowledge for policy makers. Policy makers will be able to understand real-time labor market situations in particular industries and locations. The granularity of job portal data, such as data store time stamps by minutes and seconds, and job locations by GPS code, will be useful for discussing highly precise labor market conditions, and such work can be further explored as a subsequent research to this paper. Real-time text information in job postings and resumes helps identify rapidly changing skill needs and supply, which will lead to providing demand-driven training and job placement services. The paper shows there is a great potential for Pakistan to shift its labor market policy making to a real-time data driven one as enabled by big data and technology.

References

- Amir, S., Kotikula, A., Pande, R. P., Bossavie, L. L. Y., and Khadka, U. (2018). *Female Labor Force Participation in Pakistan: What Do We Know?* World Bank.
- Atalay, E., Phongthientham, P., Sotelo, S., and Tannenbaum, D. (forthcoming). “The evolving US occupational structure.” *American Economic Journal: Applied Economics*.
- Autor, D.H., Levy, F. and Murnane, R.J., 2003. “The skill content of recent technological change: An empirical exploration.” *Quarterly Journal of Economics*, 118(4), 1279-1333.
- Beard, T. R., Ford, G. S., Saba, R. P., and Seals Jr, R. A. (2012). “Internet use and job search.” *Telecommunications Policy*, 36(4), 260-273.
- Bossavie, L., Khadka, U., and Strokova, V. 2018. *Pakistan: A Labor Market Overview (draft)*. World Bank.
- 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.
- Chowdhury, A. R., Areias, A. C., Imaizumi, S., Nomura, S., and Yamauchi, F. 2018. “Reflections of employers’ gender preferences in job ads in India: An analysis of online job portal data.” World Bank Policy Research Working Paper (No. 8379).
- Deming, D., and L. B. Kahn. 2018. “Skill requirements across firms and labor markets: Evidence from job postings for professionals.” *Journal of Labor Economics*, 36(S1), S337–S369.
- Einav, L., and Levin, J. 2014. “The data revolution and economic analysis.” *Innovation Policy and the Economy*, 14.
- Kroft, K., and 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. and Mansour, H., 2014. “Is internet job search still ineffective?” *Economic Journal*, 124(581), 1213-1233.
- Kureková, L. M., Beblavý, M., and Thum-Thysen, A. 2015. “Using online vacancies and web surveys to analyse the labour market: a methodological inquiry.” *IZA Journal of Labor Economics*, 4(1), 1–20.
- Mang, C. 2012. “Online job search and matching quality.” Ifo Working Paper (No. 147).
- Matsuda, N., and Nomura, S. 2018. “The temptation of social networks under labor search frictions.” Working paper.
- Nomura, S., Imaizumi, S., Areias, A. C., and Yamauchi, F. 2017. “Toward labor market policy 2.0: the potential for using online job-portal big data to inform labor market policies in India.” World Bank Policy Research Working Paper (No. 7966).
- Shahiri, H., and Osman, Z. (2015). “Internet job search and labor market outcome.” *International Economic Journal*, 29(1), 161-173.
- Spitz-Oener, A. 2006. “Technical change, job tasks, and rising educational demands: Looking outside the wage structure.” *Journal of Labor Economics*, 24(2), 235–270.
- Pakistan Bureau of Statistics. 2015. *Labour Force Survey 2014–15*.
- World Bank. 2010. *Stepping Up Skills: For More Jobs and Higher Productivity*.
- World Bank. 2015. *Pakistan–Enterprise Survey 2013*.
- World Bank. 2017. *Pakistan Development Update: Managing Risks for Sustained Growth*.

Appendix: Additional tables and figures

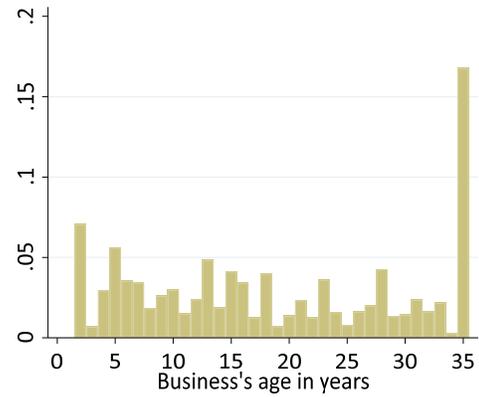
Figure A1.1. Distribution of nonagricultural enterprises by sizes (Enterprise Survey 2013)



Source: Authors' calculation using the Enterprise Survey 2013.

Note: The sampling weight is taken into account.

Figure A1.2. Age of nonagricultural enterprises (Enterprise Survey 2013)



Source: Authors' calculation using the Enterprise Survey 2013.

Note: (1) Business's age is truncated at 35. (2) The sampling weight is taken into account.

Table A1.1. Numbers of job posts and jobseekers by occupations (ISCO 2 digit)

ISCO-08	Occupation Titles	Demand	Supply
1	Managers	312,768	230,938
11	Chief executives, senior officials and legislators	8,248	19,722
12	Administrative and commercial managers	262,549	158,637
13	Production and specialised services managers	32,287	45,101
14	Hospitality, retail and other services managers	9,684	7,478
2	Professionals	439,365	309,331
21	Science and engineering professionals	80,783	65,939
22	Health professionals	10,576	11,476
23	Teaching professionals	32,975	75,857
24	Business and administration professionals	67,321	79,746
25	Information and communications technology professionals	203,798	56,392
26	Legal, social and cultural professionals	43,912	19,921
3	Technicians and associate professionals	324,564	303,325
31	Science and engineering associate professionals	14,102	30,805
32	Health associate professionals	41,343	62,262
33	Business and administration associate professionals	241,601	164,979
34	Legal, social, cultural and related associate professionals	15,814	18,649
35	Information and communications technicians	11,704	26,630
4	Clerical support workers	46,362	69,733
41	General and keyboard clerks	29,648	33,859
42	Customer services clerks	11,941	22,043
43	Numerical and material recording clerks	2,915	10,375
44	Other clerical support workers	1,858	3,456
5	Service and sales workers	118,887	75,340
51	Personal service workers	13,122	13,640
52	Sales workers	89,856	37,047
53	Personal care workers	15,896	24,633
6	Skilled agricultural, forestry and fishery workers	2,043	4,865
61	Market-oriented skilled agricultural workers	1,455	2,873
62	Market-oriented skilled forestry, fishery and hunting workers	147	1,466
63	Subsistence farmers, fishers, hunters and gatherers	441	526
7	Craft and related trades workers	19,334	30,456
71	Building and related trades workers, excluding electricians	5,524	8,079
72	Metal, machinery and related trades workers	2,390	6,040
73	Handicraft and printing workers	1,054	4,441
74	Electrical and electronic trades workers	3,932	6,330
75	Food processing, wood working, garment and other craft and related trades workers	6,434	5,566
8	Plant and machine operators, and assemblers	16,955	25,914
81	Stationary plant and machine operators	6,632	13,536
82	Assemblers	596	1,502
83	Drivers and mobile plant operators	9,727	10,876
9	Elementary occupations	21,588	15,447
91	Cleaners and helpers	1,054	2,486
92	Agricultural, forestry and fishery labourers	479	1,289
93	Labourers in mining, construction, manufacturing and transport	950	1,956
94	Food preparation assistants	3,344	1,555
95	Street and related sales and service workers	4	150
96	Refuse workers and other elementary workers	15,757	8,011

Source: Authors.

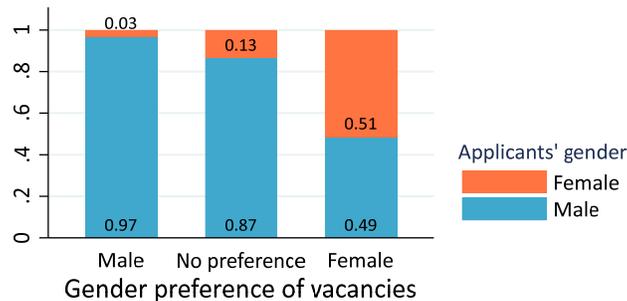
Table A1.2. Top occupations by required skill categories

Nonroutine Analytic		Nonroutine Interactive	
21 Science and engineering professionals	24.3	23 Teaching professionals	51.4
63 Subsistence farmers, fishers, hunters and gatherers	20.5	12 Administrative and commercial managers	49.1
92 Agricultural, forestry and fishery labourers	19.9	53 Personal care workers	44.0
24 Business and administration professionals	17.9	14 Hospitality, retail and other services managers	41.1
74 Electrical and electronic trades workers	17.5	61 Market-oriented skilled agricultural workers	39.3
Nonroutine Manual		Routine Cognitive	
74 Electrical and electronic trades workers	13.2	43 Numerical and material recording clerks	5.4
82 Assemblers	12.7	63 Subsistence farmers, fishers, hunters and gatherers	5.3
73 Handicraft and printing workers	6.7	24 Business and administration professionals	4.0
72 Metal, machinery and related trades workers	5.4	44 Other clerical support workers	3.7
31 Science and engineering associate professionals	4.8	93 Labourers in mining, construction, manufacturing and transport	3.6
Routine Manual		Cognitive	
82 Assemblers	10.9	41 General and keyboard clerks	25.6
81 Stationary plant and machine operators	8.7	92 Agricultural, forestry and fishery labourers	16.2
74 Electrical and electronic trades workers	7.4	63 Subsistence farmers, fishers, hunters and gatherers	9.8
31 Science and engineering associate professionals	6.5	42 Customer services clerks	9.5
73 Handicraft and printing workers	5.9	23 Teaching professionals	8.9
Social		Character	
96 Refuse workers and other elementary workers	15.5	96 Refuse workers and other elementary workers	34.7
93 Labourers in mining, construction, manufacturing and transport	15.3	53 Personal care workers	25.2
33 Business and administration associate professionals	13.5	93 Labourers in mining, construction, manufacturing and transport	22.0
24 Business and administration professionals	12.2	24 Business and administration professionals	21.9
52 Sales workers	12.2	74 Electrical and electronic trades workers	21.8
Writing		Customer Service	
26 Legal, social and cultural professionals	21.0	12 Administrative and commercial managers	45.0
93 Labourers in mining, construction, manufacturing and transport	3.9	14 Hospitality, retail and other services managers	36.8
73 Handicraft and printing workers	3.2	52 Sales workers	35.3
24 Business and administration professionals	3.1	33 Business and administration associate professionals	35.0
25 Information and communications technology professionals	3.0	73 Handicraft and printing workers	30.5
Project Management		People Management	
63 Subsistence farmers, fishers, hunters and gatherers	20.7	14 Hospitality, retail and other services managers	20.5
96 Refuse workers and other elementary workers	14.0	96 Refuse workers and other elementary workers	19.9
61 Market-oriented skilled agricultural workers	13.1	91 Cleaners and helpers	18.8
13 Production and specialised services managers	13.1	92 Agricultural, forestry and fishery labourers	18.6
31 Science and engineering associate professionals	12.6	44 Other clerical support workers	16.8
Financial		Computer	
24 Business and administration professionals	10.3	35 Information and communications technicians	16.5
43 Numerical and material recording clerks	9.4	41 General and keyboard clerks	14.7
13 Production and specialised services managers	7.6	92 Agricultural, forestry and fishery labourers	12.7
12 Administrative and commercial managers	5.9	25 Information and communications technology professionals	12.6
44 Other clerical support workers	5.7	44 Other clerical support workers	9.5

Source: Authors.

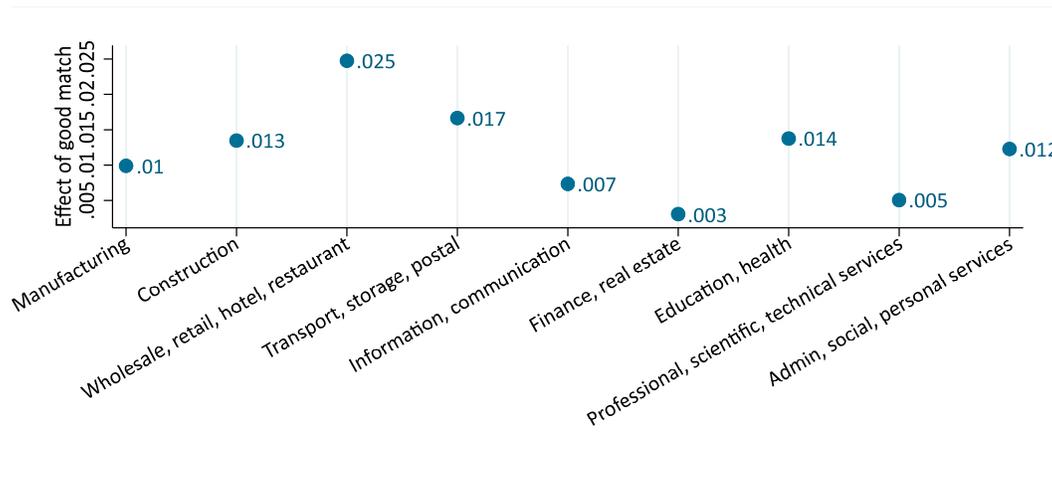
Note: Shown is the frequency of task-related words per 1,000 job description words.

Figure A1.3. Types of applicants by gender



Source: Authors' calculation using the data from Rozee.pk.

Figure A1.4. Industry in which a good match between applicants and jobs matter



Source: Authors.

Note: Shown is the coefficient of dummy for good industry match by industries.