Bridging the Miles: Spatial Factors in Job Application and Selection*

Kanika Mahajan† Shekhar Tomar‡

Ashoka University *ISB*

August 17, 2023

Abstract

Using data from a large online job portal on posted jobs and applications and accounting for unobserved heterogeneity at the job and candidate level, we find that candidates are 90% less likely to apply to jobs beyond 100 miles in India. Albeit, the observed distaste is higher towards jobs located in smaller non-metro cities. To examine the channels, we first exploit a natural experiment that introduced subsidized air linkages between cities. While such a policy change increases applications to distant metro locations by 25%, there is little effect on applications for jobs located in distant non-metros. We also find that higher availability of jobs in a city within the candidate's target occupation substantially offsets this observed distaste towards jobs in non-metros, underlining the influential role of prospective job mobility concerns in explaining distaste for distant locations. Utilizing the shortlisting information, we finally rule out that applicants' behavior is an equilibrium response to employer behavior. Employers in non-metro cities show no discernible distaste to shortlisting candidates from distant locations.

JEL Codes: J61, J20, J16, E24, R23

Keywords: Job search, Distaste for distance, Air transport, Job Concentration

*We thank Nikhil Datta, Robert Garlick, Krishna B. Kumar, and conference participants at Jobs and Development Conference (Cape Town, 2022) for helpful comments and suggestions. Vandana Saini provided excellent research assistance. We thank the private job portal for providing access to their data. All remaining errors are ours.

†Ashoka University, Haryana, India. Email: kanika.mahajan@ashoka.edu.in

‡Indian School of Business, Mohali, India. Email: stomar.dse@gmail.com

1 Introduction

Frictions arising from low mobility of labor across locations is often seen as one of the major sources of geographic mismatch in the labor markets (Porcher, 2020; Marinescu & Rathelot, 2018). At the same time, larger cities tend to attract more migrants (Fujiwara *et al.*, 2022; Yakubenko, 2020), which may be attributed to better information about jobs, higher growth of jobs in larger cities, better amenities, or the presence of skill complementarities. Consequently, this affects spatial sorting and agglomeration. A few questions, however, remain unanswered. First, conditional on information availability, to what extent do job seekers prefer larger cities? Second, can lowering transport connectivity costs induce job seekers to sort differentially across cities? Finally, do future job mobility and employer preferences play any role in generating the observed sorting across cities.

The importance of these questions becomes particularly pronounced in the context of developing countries where spatial growth is typically more concentrated (Chauvin *et al.*, 2017). Notably, India, the backdrop for our investigation, presents a compelling setting for studying spatial concentration resulting from limited migration over long distances between its numerous cities. In a sharp contrast to countries like the United States, service sector growth in India has primarily been concentrated in areas that were high-density to begin with (Desmet *et al.*, 2015). This fact is underscored by the 2011 Census data, which shows that the number of working age migrants who moved to larger metropolitan cities for employment during the decade 2001-2010 as a ratio of their base population was 4.7 percent. This proportion was 3.8 percent for other cities. Whether place-based policies can help reshape the spatial distribution of employment remains an open policy question.

We analyze candidate job search behavior over distance using data from approximately 37,000 job ads posted for hiring full-time employees, and the 226,000 candidates who applied to these vacancies through a large online job portal in India, between July–December 2020. Primarily, the portal caters to young, first-time, urban job seekers who are likely to be relatively more mobile. We observe job characteristics, like education and experience requirements for almost all jobs, and posted wages (for a majority of jobs). On the candidate side, we observe attributes like age, gender, education, experience, city of current location, and most importantly, all the job applications made by them to the aforementioned ads.

We first estimate the overall distaste for distance by job seekers by examining their applications across cities. We construct a choice set for each candidate based on the job role and date of application. It gives us around 166 million observations at candidate×job ad level, with an average application rate of 1 percent.[1] The rate of application is the highest at 6.5 percent within a job seeker's own city. Using the above choice set and accounting for job and candidate fixed effects, we find that the probability of application falls as the distance between candidate and job location increases. Quantitatively, the application rate for jobs located at least 100 miles away, relative to jobs in own city, falls by 90 percent. In fact, the application rate approaches zero for a distance above 500 miles.[2] We find that candidate attributes, like gender, education, and experience, influence the distaste but play a small role in explaining its magnitude.

We next examine whether the application probability differs for jobs located in large metropolitan (metro) cities vs. other cities. We divide the cities in India into metro and non-metro locations, based on the population of their urban agglomerations as reported in the latest Indian Census of 2011.[3] Delhi, Mumbai, Bengaluru, Chennai, and Kolkata are the top five metro cities based on this classification. These metro cities have better amenities like healthcare, education, and inter-connectivity with the rest of the country. If such amenities matter, candidates would be more inclined to apply to jobs in metro cities

[1]This rate is close to the average obtained experimentally by He *et al.* (2021) for a job board in China at 0.5%.

[2]The magnitude for distaste to distance in our sample is intermediate. It is lower than that for the U.S. (Marinescu & Rathelot, 2018) and higher than that for Chile (Banfi *et al.*, 2019), the two studies which estimate the distaste using applications data. We also find a 37% lower distaste for distance by more educated candidates (similar to Marinescu & Rathelot (2018) and Wozniak (2010)), while more experienced candidates have a higher distaste by 10%. Females are less likely to apply to distant jobs, and their additional distaste is 9.4% of the average distaste for distance beyond 100 miles.

[3]For a list of all cities by urban population, see Census 2011. Four of these cities have constituted major metropolitan areas in India and have been the focal point of increasing road connectivity (See: Indian Express).

relative to equidistant non-metro cities. Our analyses validate this hypothesis, as the distaste for distance (beyond 100 miles) for non-metro cities is at least twice (or 200 percent higher) than that for metro cities. Above-median wages reduces this additional distaste for non-metro jobs by 38 percent (above 100 miles), suggesting that the disutility of working in non-metro cities can be partially offset by higher wages. These findings are robust to including overall distaste by candidates for non-metro jobs and also by comparing job ads within a firm-job role but located in different cities. We find that the distaste for jobs in farther non-metro locations holds despite adding these stringent controls. We also use data from 2019 on posted jobs and applications made by candidates, to account for any effect of the pandemic on the distaste for distance, and find similar results.[4]

To understand the distaste for non-metro locations, we explore several competing channels, hitherto undocumented in the literature. First, we test whether decreasing connectivity costs can increase the applications by job-seekers by exploiting a natural experiment that increased subsidized air connectivity across Indian cities under a scheme called *UDAN*. Implementing a staggered difference-in-differences estimation strategy, we find that increased connectivity leads to 25 percent higher applications towards jobs in metro locations, yet it doesn't significantly affect applications for non-metros. On average, larger cities have better connectivity vs. smaller cities. If such connectivity costs matter, then a reduction in these costs can increase mobility of job-seekers. However, if other costs or factors matter more significantly then such place-based policies may not have any effect on job applications.

Next, we consider other factors that could potentially contribute to the pronounced distaste for similar jobs in non-metro locations. These include concentration of jobs in the city in the same occupation, which affects future job mobility, and cultural and language proximity. We use the average concentration of occupation or job role in a city as well as exploit the timing of job search spell to calculate the proportion of jobs posted across cities in a given job role visible to a given candidate on the portal. We find that a higher proportion

[4]We use the estimates from 2020 in our main analyses to maintain consistency since the shortlisting data for employers is only available for this period.

of jobs posted for the same job role decreases the distaste for distant cities. Remarkably, for the non-metro locations the distaste almost completely disappears beyond 100 miles if the city has 50 percent postings for a given job role in a search spell. However, cultural and language proximity do not explain the lower preference for distant non-metro locations.

Lastly, we test whether the candidate distaste is an equilibrium outcome where employers exhibit similar distaste for non-local candidates. For this, we use data on candidate shortlisting by employers, which is available for a subset of 1,470 job ads where the employers purchased this service from the portal. Accounting for job and candidate level unobservables through fixed effects, we find that the shortlisting probability only diminishes by 14 percent for candidates located at least 500 miles away from the job location, with no differential effect for smaller distances. Notably, employers hiring for metro-based jobs are 26 percent less likely to shortlist candidates situated over 500 miles away, whereas those hiring for non-metro jobs are 11 percentage points more likely to consider such candidates, effectively offsetting the baseline distaste for distant applicants in non-metro jobs. Taken together, these results show that the large distaste for distance by job seekers is mostly explained by job seeker preferences rather than the job seekers facing a lower probability of selection in distant non-metro jobs.

Our work contributes to three broad strands of literature. First, it is related to the literature examining factors that explain spatial sorting in the labor markets. Apart from information, better amenities in larger cities like better living conditions in terms of consumption goods and services (Henderson, 1986; Ades & Glaeser, 1995), better human capital growth (Glaeser & Resseger, 2010; Glaeser & Mare, 2001), skill complementarities (Eeckhout *et al.*, 2014) and other amenities (Albouy *et al.*, 2020; Albouy & Stuart, 2020; Carlino & Saiz, 2019; Clark *et al.*, 2002; **?**) are shown to be positively related with higher population shares or growth in larger cities.[5] We extend this literature by studying how lower connectivity costs also increase spatial mobility by affecting job seeker preferences for distant locations.

[5]On the other hand, preference for large cities can also be lower due to dis-amenities like pollution and congestion, but (Gollin *et al.*, 2021) find little effect of these contributing to the wage premium in larger cities.

However, the benefit can be disproportionate for larger and smaller cities, with larger cities attracting more job seekers vis-a-vis smaller cities. We also find evidence for job concentration in cities, that increases future job mobility in case of job separation, as another possible mechanism.

We directly contribute to the literature on connectivity costs and its impacts. Existing studies have estimated the economic impacts of transportation infrastructure projects like highways and railways on employment or urban growth (Duranton & Turner, 2012; Lin, 2017), urban forms (Baum-Snow, 2007), skill premiums (Michaels, 2008), long-term economic growth (Banerjee *et al.*, 2020) and trade (Donaldson, 2018).[6] The impacts of increased air connectivity are relatively less evaluated. For instance, Blonigen & Cristea (2015) and McGraw (2020) find that expansion of airports in the US increased economic and population growth in cities that were connected while Sheard (2014) find no impact using a 1944 policy change in the US. Gibbons & Wu (2020) find a positive effect on manufacturing activity due to an expansion of airports in China. Breidenbach (2020) find no impact of regional airports in Germany while Zhang & Xie (2023) find heterogeneous impacts across Chinese cities with those having higher economic activity initially gaining more. We extend this literature by directly showing the impact of reducing connectivity costs on applications by job seekers. While economic and employment growth can be impacted due to increase in job creation as connectivity costs fall, our study is the first to provide the direct impact on relocation preferences of job seekers.

Third, our study is related to the work that examines the location preferences of job seekers across geographic locations. Marinescu & Rathelot (2018) using aggregate application data across the U.S. zip codes, find that application rates fall by almost 100% beyond 50 miles. Banfi *et al.* (2019) using individual application data from a Chilean job portal find that the application probability falls by at most 75% for jobs located beyond 200 miles. Our work contributes to this literature by showing the heterogeneous distaste by city size. Our findings,

[6] Apart from direct labor market and economic impacts studies also examine the impact on innovation, entrepreneurship, information acquisition in lending and pollution.

thus, show that smaller locations face higher costs of geographic mismatch. Lastly, this is the first paper to document employer distaste for distance in shortlisting candidates. While statistically significant, this distaste is one-sixth of the magnitude of the candidates' distaste for distance and shows that the latter cannot be explained by employer discrimination.[7] Furthermore, our paper is the first to provide evidence of distaste for distance in a developing country context.[8]

More generally, our paper is related to the literature that looks into the job attributes valued by job seekers (Chaturvedi *et al.*, 2021; Rickne & Folke, 2020; Mas & Pallais, 2017; Wiswall & Zafar, 2018; Eriksson & Kristensen, 2014). It is also related to a growing literature using data from online job portals to examine various aspects of labor market search. For instance, studies examine how search by job seekers is affected by search duration, (Kudlyak *et al.*, 2012; Faberman & Kudlyak, 2019); gender differences in job search (Kuhn & Shen, 2021; Chaturvedi *et al.*, 2021; Fluchtmann *et al.*, 2021; Kuhn *et al.*, 2020) and applications over business cycles (Ma & de la Parra, 2021; Hensvik *et al.*, 2021). It shows that data generated through online matching platforms can also be useful for evaluating impacts of changes in public policy on candidate search behavior.

The remainder of the paper is organized as follows. Section 2 discusses the data and Section 3 elucidates the empirical strategy for measuring candidate or job seeker distaste over distant jobs. Section 4 discusses the findings while Section 5 evaluate the channels that contribute towards the observed higher distaste in applying to distant non-metro cities

[7] Le Barbanchon *et al.* (2021) using French administrative data for unemployed workers find that women value commute 20% more than men. Further, using final hiring data they show that while men and women both are less likely to be hired if located farther within a city, women applicants face no additional disadvantage. To the extent that final hiring may be affected by both candidate and employer preference, it is difficult to disentangle between the two. Phillips (2020) overcomes this through a correspondence experiment and finds that employers in Washington D.C. are less likely to shortlist candidates who live 5-6 miles away. It does not examine impacts by gender. Our paper is the first to conduct employer-level analyses using data on applications and shortlisting for jobs located in distant geographical location rather than in a within city context. We find that employers are less likely to shortlist applicants when they are located at least 500 miles away, driven by female applicants, and jobs in metro locations where the applications per vacancy on average are larger. Thus, showing that employer distaste stems from perceived costs of shortlisting candidates.

[8] Geographical mobility can be lower in low-income countries due to credit constraints. It can also be higher due to a lack of opportunities in a candidate's location.

including connectivity costs, future job mobility and employer distaste. Section 6 concludes.

2 Data

Our data comes from one of the major online job platforms that primarily caters to young job-seekers in India. The platform allows candidates to create their profiles and apply for positions at no cost, while employers can post job advertisements for a nominal fee of USD 20 per job ad, with a validity of up to 60 days, and receive applications. The portal also offers premium services to employers for an additional fee, enabling them to shortlist candidates and communicate with them through their interface.[9] The final hiring decision, however, is not observed by the portal.

The primary analyses are drawn from data related to job ads posted from July–December 2020, and the applications made to these jobs. For a subset of these jobs ads, data on shortlisting by the employers, premium service subscribers, is also available. Thus, we observe information on both sides of the matching market during this time period.[10] Additionally, we have data on posted ads and applications from June 2018–December 2019. We use it for robustness checks and supplemental analyses in Section 5.

Our first dataset pertains to posted jobs, where we have extensive information such as date of posting, job role, job title, education requirements of the job ad, minimum and maximum required experience, minimum and maximum offered wage, job location, the firm posting the job ad, and other characteristics like number of vacancies. There were 51,549 full-time job ads posted on the platform during July 2020-December 2020. We restrict our sample to full-time job ads that mention a single city for job location, receive at least one candidate application, and provide a specific job role rather than "Others". After removing ads that have no applications within the first ten days of posting, we are left with a final sample of 37,045 job ads across 419 cities. We elaborate the importance of this last step in

[9]Premium services cost USD 70 for sending upto 1200 call letters and USD 140 for upto 4000 call letters.

[10]This period does not overlap with the COVID-19 induced national lockdowns in India. The first national lockdown was imposed on 24 March, 2020 and most restrictions removed by May 31, 2020.

the next section.

On average, each job ad receives 45 applications. We categorize cities into metro (Delhi, Mumbai, Chennai, Kolkata, and Bengaluru), and non-metro (all others). The metro cities, accounting for 60 percent of the jobs in our dataset, are significant economic and population hubs. For instance, Delhi and Mumbai are the largest cities in north and west India, respectively. Appendix Figure A.1 shows the distribution of cities across the country for the posted job ads. We provide descriptive statistics for these jobs ads in Appendix Table A.1. As mentioned earlier, these jobs largely cater to young job seekers who are inexperienced, with 50% and 33% jobs posting a requirement of 0-1 and 1-2 years of experience, respectively. Notably, about 56% of the jobs ads post wages, which is larger than many other existing studies that use posted wages.[11]

The second dataset provide information on candidate characteristics, like age, gender, education, experience and most importantly all the job applications made by them to the posted ads. After excluding candidates with incomplete profiles, we are left with 0.68 million applications made by 226,464 candidates located across 679 cities (Appendix Figure A.2).[12] Approximately 44% of them reside in metro locations. We provide descriptive statistics for candidates in Appendix Table A.2. On average, these candidates apply to 4 jobs in our job ads sample during this time period.

Finally, the third dataset provides information on shortlisting. We observe which candidates are shortlisted by the firms for a given posted job in case employers use the premium service provided by the portal. Shortlisting is observed for 1,470 job ads. To be consistent with the previous datasets, we keep the 888 job ads that do not mention "Others" in their job

[11]For instance, wages are advertised in just 13.4% in Banfi *et al.* (2019) and Banfi & Villena-Roldan (2019), 16.4% of job ads in Kuhn & Shen (2013), 20% of job ads in Marinescu & Wolthoff (2020) for the U.S. and 24.8% of job ads in Brenčič (2012). In line with the existing studies, we also find that wages are less likely to be posted for ads having higher skill requirements like higher education and experience, thus, allowing for possibility of wage bargaining and negotiation in higher skill jobs (Brenčič, 2012; Banfi & Villena-Roldan, 2019; Michelacci & Suarez, 2006).

[12]We drop candidates with incomplete profiles, specifically the ones who do not fill up gender information. We find that candidates who do not mention gender, do not enter other characteristics like education and date of birth either. They also make much fewer applications (less than 1/10th of those made by candidates who reveal gender). We drop these non-serious candidates applicants from our analyses.

role. Around 83% of these jobs are in metro locations, almost all specify wage and on average have 63 applicants per job ad. Appendix Table A.3 gives summary statistics for these jobs.

More comprehensive information regarding these datasets, including the specifics of data processing, are detailed in Appendix Section A.

3 Empirical Strategy

In this section, we discuss our main estimation strategy to examine the distaste of candidates for distant jobs and how it is mediated by other job or candidate level characteristics. We begin by describing the construction of job choice sets for candidates.

3.1 Construction: Job Choice Sets for Candidates

Our principal empirical objective is to investigate how the geographical distance between the location of job postings and the location of candidates influences application rates. While we observe the set of job ads that candidates apply to, we lack information on the jobs that they choose to not apply to. Consequently, we need to construct choice sets of potential jobs for these candidates for our analyses.[13] For a given candidate i applying to job role r on date d, we include all jobs postings on the platform in job role r between $d-10$ and $d+1$ days in her choice set. This construction is informed by two main features in the data.

First, the job search function on the website lists job ads by their posting date as default. More recent job ads are listed at the top and have a higher likelihood of being seen by a candidate, which is corroborated by the data. Figure 1, Panel (a) plots the average number of applications received by a job ad from its posting date. The x-axis shows the number of days elapsed since the date of posting. The number of applications a job ad receives is

[13] Instead of constructing choice sets, an alternative approach could involve estimating aggregate application probabilities by determining the likelihood that an applicant from city i applies to jobs in city j (Marinescu & Rathelot, 2018). While such an approach is useful at an aggregate level, it does not allow to control for candidate- and job-specific variation. Moreover, constructing a choice set enables an examination of differential behaviors in response to job and candidate attributes, as we discuss later.

highest during the first two days (7 per day) and then falls steeply. By the end of 10 days, job ads receive around 2 applications per day. The initial 10 days account for 60 percent applications received by a given job ad, which justifies the $d-10$ cutoff.

Second, candidates bunch their applications around the date of their first application. Panel (b), presents the average number of applications made by candidates after their first appearance in the data, i.e. after their first application. On day 0, candidates make an average of 2.4 applications. This number sharply drops to 0.2 on day 1 and approaches zero thereafter, rationalizing the $d+1$ cutoff in our choice set construction. These patterns in search behavior are consistent with findings in (Davis & de la Parra, 2017) for the US, suggesting that candidates apply to jobs over a very short period of time, and become inactive thereafter.

Overall, our approach yields a choice set comprising 82 million candidates×job observations, with a mean application rate of 0.48 percent. In the main paper, we present results using the above choice set, but our findings are robust to employing different cut-offs in the choice set construction process.[14]

3.2 Measuring Candidate Distaste over Distance

Using the above choice set, we measure the candidate distaste for distant jobs by using the below specification:

$$Applied_{ijkt} = \beta_0 + \sum_{k=1}^{g} \beta_k DistGroup_k + \alpha_i \mathbb{X}'_i + \alpha_j \mathbb{X}'_j + \delta_i + \delta_j + \delta_t + \epsilon_{ijkt}. \tag{1}$$

$Applied_{ijkt}$ is an indicator variable that takes a value of one if candidate i applies to job posting j located at a distance k in period t, else zero. Our main explanatory variables

[14] In general, the problem of constructing consideration sets or in our case arriving at the set of jobs which were considered by the candidate before making an application, is pertinent in the industrial organization literature (Van Nierop *et al.*, 2010). Generally, consideration sets can be created using common attributes. For instance, Le Barbanchon *et al.* (2021) use geography, occupation and time horizon to construct the choice set. Banfi *et al.* (2019), on the other hand, use networks of job seekers who make common job applications to arrive at the set. If candidate A and B both apply job i but candidate A applies to j too, then j will form a part of the consideration set for Candidate B, in this approach.

include $DistGroup_k$, denoting the distance between the candidate's city i and the city of job location j. The variable $DistGroup_k \in \{1-50, 50-100, 100-500, 500+\}$ miles and takes a value of one if the distance between the candidate–job location pair lies in the k^{th} distance bracket, else zero.[15] The reference group in this regression comprises jobs located at zero distance from the candidate i.e., jobs located within the candidate's home city. Our main coefficient of interest β_k measures the difference in application likelihood with the distance between location i and j, relative to a job located in the candidate's city of residence. If application likelihood decreases with distance, then the coefficients β_k would become more negative for higher values of k.

We account for any variation coming from job-level controls ($\mathbb{X}_j$) like city, qualification, required experience as well as candidate-level controls ($\mathbb{X}_i$) like gender, education and age. The specification also includes candidate (δ_i), job (δ_j), and month-year (δ_t) fixed-effects. δ_i filters out any unobserved candidate-level factors, δ_j controls for any unobserved job-level characteristics, and δ_t accounts for time-varying factors common to all candidates that could affect application. Our identification, therefore, comes from comparing the application behavior of a given applicant i across jobs j that she could potentially apply to. In the most saturated specification, the coefficients on $\mathbb{X}_i$ and $\mathbb{X}_j$ cannot be estimated as the non-time varying factors are absorbed in δ_i and δ_j. We weight the regressions by the inverse of total applications made by a candidate in order to give equal weight to all candidate-applications in our regressions. Otherwise, candidates who make more applications would be assigned a larger weight in the specification. The standard errors are clustered at the candidate level.

We augment the baseline Equation 1 to estimate if distaste to distance varies across

[15] We use geodist command in Stata to calculate the distance between two locations. We also estimate specifications with more disaggregate groups but eventually use this distance grouping as coefficients are similar for groups beyond 100 miles.

candidate or job-level characteristics, by using the following specification:

$$Applied_{ijkt} = \beta_0 + \sum_{k=1}^{g} \beta_k DistGroup_k + \sum_{k=1}^{g} \beta'_k (DistGroup_k \times \mathbb{C}) + \alpha_i \mathbb{X}'_i + \alpha_j \mathbb{X}'_j + \delta_i + \delta_j + \delta_t + \epsilon_{ijkt} \tag{2}$$

where $\mathbb{C}$ represents the characteristic under consideration and β'_k is the coefficient of interest on the interaction term $DistGroup_k \times \mathbb{C}$. One critical source of heterogeneity that we estimate pertains to the difference in distaste for jobs located in metro vs. non-metro cities. In this case, $\mathbb{C}$ is a dummy variable assigned a value equal to one for non-metro cities, and zero otherwise. A negative β'_k would suggest that candidates are less likely to apply to jobs in a non-metro city vs. metro city, even when the two jobs are equidistant.

4 Candidate Distaste for Distant Jobs

In this section, we present our principal findings regarding the variability in application rates contingent on the distance between the candidate-job location city pair and how it varies depending on where the job is located, metro vs. non-metro.

4.1 Overall Distaste for Distance

Table 1 reports the results on candidate distaste by estimating Equation 1. Each column exhibits variations corresponding to the distinct set of controls and fixed-effects incorporated in the regression. Column (1) includes month-year fixed-effects to account for variation in application rates over the time period. We find that the probability of application goes down as the job location becomes more distant relative to candidate's home city. The coefficient associated with each k-th distance group are negative and significant, with the likelihood of application declining by around 6 percentage points for jobs located $100 - 500$ or 500+ miles away from the candidate's city. Considering the baseline application probability within the candidate's city is 6.5 percent, these coefficients suggest a near-zero probability of application for jobs more than 100 miles away.

Alternate specifications (Columns (2)–(4)) yield results of similar magnitude. Column (2) includes candidate-level controls and job fixed-effects, thereby mitigating the effect of unobserved job-level heterogeneity. Column (3) includes job-level controls with candidate fixed-effects, accounting for unobserved candidate-level variance. Finally, column (4) includes both job and candidate fixed-effects and corresponds to the most saturated specification accounting for any unobservable factors that can influence application rates based on job and candidate attributes. Consistent with column (1), the fall in application rate for jobs located a minimum of 100 miles away, relative to a candidate's home city, exceeds 90 percent.

The magnitude of distaste among the Indian job seekers is intermediate when compared to other papers in the literature. Banfi *et al.* (2019) report that amongst young unemployed job seekers in Chile, the application rate declines by 40 percent for jobs located more than 200 miles away. On the other extreme, Marinescu & Rathelot (2018) find that the probability of application falls to almost zero beyond 75 miles in the U.S. Our findings reinforce the distaste for distance as a robust feature of intercity job search behavior as our specifications represent a significant advancement over previous studies. We find that distaste result is robust and withstands an extensive battery of candidate- and job-level fixed-effects, hitherto not tested in the literature.

We consider the role played by various factors in influencing the candidate distaste for distance in Appendix Section A.1. For instance, if candidates experience a disutility associated with relocation to a distant location, higher wages could potentially serve as a compensating factor. For this examination, we focus on the subset of jobs featuring posted wages.[16] Appendix Table B.1 presents the findings pertaining to the impact of wages on distaste. We find that high wages leads to a modest reduction in distaste for distance. We

[16]Around 56% of the job ads report wages in our data and we construct the potential choice set over these jobs for each candidate. This yeilds around 62 million candidate$\times$job observations with an average application rate of 1.7%.Jobs with posted wages are likely to see a higher application, even after controlling for other job-level characteristics and candidate fixed-effects. We do not report these results in the main paper, but they are available on request.

report similar heterogeneity in distaste by gender, education and experience of the candidate in Appendix Table B.2.

These results reveal that while posted wages and other candidate attributes indeed influence application probability and the distaste towards distant jobs, they do not quantitatively account for the candidate's disinclination to apply for jobs that are located far away. This leads us to investigate the next key attribute, city type, in generating distaste for distance among candidates.

4.2 Distaste for Non-metro Locations

In this subsection, we explore whether a candidate's preference for metro cities contributes substantively to the documented distaste for distant job opportunities. Metros, by virtue of their superior amenities, might appeal more to potential job seekers than non-metro locations. Features such as better connectivity and infrastructure, a wider range of job opportunities fostering better mobility within job roles, access to social networks for candidates who may have local acquaintances in these cities, as well as enhanced language or cultural proximity may generate more preference for metro cities.

Table 2 reports the differential distaste between jobs posted in metro vs. non-metro cities. We estimate Equation 2 where we incorporate interactions between an indicator variable that takes a value of one for jobs located in non-metro locations, zero otherwise, and the distance groups. Column (1) reports the results accounting for job fixed-effects, candidate fixed-effects and month-year time fixed-effects. We find that the coefficients on k-th distance groups as well as their interaction with non-metro city are negative and significant. Quantitatively, the likelihood of application declines by an additional 7.4 percentage point for non-metro jobs located 500 miles away from the candidate. Therefore, the average distaste for non-metro jobs is almost triple than that for a metro city job located 500 miles away. While jobs located in distant metro cities witness distaste, the degree is insufficient to completely nullify the average probability of application.

In column (2), we estimate the distaste across metro vs non-metro cities for jobs that post wages in our data. The application rate is higher at 11% for these jobs, leading to higher magnitudes for distaste in percentage points. We find that candidates are 7.5 percentage point or 70 percent less likely to apply jobs in metro cities located 500 miles away. This distaste nearly doubles (by 15 percentage points) for non-metro locations.

It is plausible that the above results for non-metro jobs are on account of lower wages in non-metro jobs rather than the value of city-level amenities to job seekers. To test this hypothesis, we estimate the same specification as in column (2) adding an additional control that interacts the distance groups with an indicator variable for the posted wage being above the median posted wage (column (3)).[17] The higher distaste for non-metro jobs persists in this specification. Lastly, in column (4), we interact the indicator variable for non-metro job with that of a job posting above median wage (distance×above median wage×non-metro). Here, we find that the compensating effect of wages primarily stems from non-metro jobs located father away. A non-metro job posting a higher than median wage results in decreasing the distaste by 7 percentage points or 38 percent(=0.07/(0.184)) when a job is located more than 100 miles away. On the other hand, the reduction in distaste for metro jobs is 7.7 percent when jobs post a higher than median wage.

4.3 Robustness

The above results are robust to alternate sample of candidates, additional controls, and alternate periods.

Sample of College-educated Job-seekers: We examine the effect on candidates having at least college education as the distaste for distance is marginally lower for these candidates. We find that they have similar distaste for non-metro locations (Appendix Table B.3).

[17] We also estimate a specification where we interact log wage with distance groups, and find similar results. We omit these for brevity but are available on request.

Candidate or Job-specific Controls: Appendix Table B.4 rules out that non-metro distaste results are not driven by candidate-level heterogeneity. For a given candidate, we compare the distaste across non-metro locations based on distance by including an interaction term of candidate FE with an indicator variable for non-metro locations in column (1). The distaste for more distant non-metro locations remains higher than metro ones. Column (2) rules out any role played by differential application rates across jobs due to their education and experience requirements. Lastly, in column (3) we interact candidate FE with more than median wage and non-metro location, when examining whether posting higher wages reduces the distaste. We continue to find that the distaste for distance reduces by almost 40 percent for farther non-metro locations when they post an above median wage.

Non-metro Distaste – Heterogeneity Tests: While our baseline regressions include job fixed-effects, occupations and firms in non-metro locations might significantly differ from those in metro. Our findings remain robust after including various interactions of job characteristics with distance, as shown in Appendix Table B.5. We control for differential distaste over distance by education and experience (column 1), job role (column 2), detailed occupation based on job title (column 3) and detailed occupation by posting company (column 4). The last regression represents a stringent specification as we compare jobs posted by the same firm within the same occupation (granular job title level), yet we continue to observe a significant distaste for distant non-metro locations.

Alternate Periods: Our primary analyses uses data from the latter half of 2020. However, our results are robust to using data from alternate periods. We report results based on October–November 2019, creating a choice set for this period and estimating the overall distaste for distant jobs and the additional distaste for non-metro locations. Appendix table B.6 reports the overall distaste for distance. Appendix Table B.7 confirms that distaste is higher for non-metro locations, aligning with results reported in Table 2.

To summarize, we observe a precipitous decline in the application probability as the distance to job location increases. A considerable fraction of this decline can be attributed to non-metro cities located far away from candidates' location.

5 Non-metro: Which Attributes Matter for Distaste?

This section investigates the major factors that contribute to the reduced likelihood of candidates applying to jobs in distant non-metro cities.

5.1 Long Distance Travel Connectivity

One of the factors why candidates could prefer metro cities is the ease of connectivity from metro to other locations. The effect of such connectivity on job application would be more pronounced if the job city is located far away from the candidate's own city. We examine whether policies aiming to reduce long-distance travel costs (both in terms of time and money) can impact candidate's preference to apply to distant non-metro jobs.[18]

We exploit the natural experiment setting provided by *UDAN* scheme that enhances air travel connectivity from small cities to other parts in India. The schemes introduces new air routes at subsidized airfare, approximately half the cost of road travel. Airlines bid for air routes, with the contract being awarded to the one requiring the least subsidy.[19] The first flight under the scheme was launched in April 2017.

Our application data from June 2018 – December 2019 helps us to estimate if improved long-distance connectivity impacts candidate applications to cities that get a route under the *UDAN* scheme. If the scheme reduces travel time and cost from candidate's home city to job

[18] There is a broad literature studying the economic impact of transportation infrastructure. Among these, Lin (2017) shows how High Speed Railway connectivity in China boosted city-wide passenger flows and employment. Our paper is the first one to provide direct evidence of enhanced connectivity on candidates' job application behavior.

[19] Under the scheme, the airfare for a one-hour journey by a 'fixed wing aircraft' or half an hour's journey by a helicopter for about 500 km, was been fixed at INR 2500 (USD 75) at the time of launch. Aviation firms operating these routes are provided funds under the Viability Gap Funding to cover their losses due to low airfares. Airport fees and other charges are also waived off and electricity and other services are provided free of cost to the airlines. Currently, traveling by taxi in the country costs an average of Rs 10 per kilometer.

location, it may influence candidates' decision to relocate to distant locations. Therefore, the scheme could potentially reduce distaste for distance by minimizing the cost of relocation and future travel expenses to home city. This scheme offers a pertinent setting to gauge how connectivity might contribute to distaste for non-metro locations.

We use the data for cities where candidates are located and which receive an UDAN airplane route until June 2022. We exclude larger cities with pre-existing airports, as the scheme would have minimal impact on connectivity of these cities.[20] This leaves us with 54 cities where candidates are located. On the jobs side, we keep the cities which receive at least one application from any of these 54 cities and also get connected by at least one route under *UDAN*.[21] Given that airport connectivity is likely to be a significant factor only over medium-to-long distances, we limit the city pairs to those that are at least 200 miles apart. This leaves us with a final sample of 83 routes that that got connected under the scheme.

To estimate the impact of this scheme on candidate applications, we compare the applications received by city k from candidates whose home cities get connected through *UDAN* (treated groups) vs. those cities that do not get connected (control group), before and after the opening of route. Since the routes open in a staggered fashion, we use the estimation methods from the growing difference-in-differences (DID) literature that allows to estimate the average treatment effect after taking into account the timing of treatment (Goodman-Bacon, 2021; Callaway & Sant'Anna, 2021; Sun & Abraham, 2021).[22] We estimate the following

[20] These cities include - Ahmedabad, Amritsar, Bangalore, Bhubaneswar, Chandigarh, Chennai, Delhi/NCR, Dibrugarh, Hyderabad, Imphal, Jaipur, Kochi, Kolkata, Lucknow, Mumbai, Patna, Pune, Thiruvananthapuram, and Varanasi.

[21] Candidates from 54 cities are connected to new locations under 127 routes. For our main estimation, we consider routes that were launched after June 2018 since the applications data begins from June 2018. We show robustness to including routes that receive treatment earlier. Also, we keep those job location cities that receive at least one route under *UDAN*. This sample of cities with some airport infrastructure is comparable to each other, relative to other cities in the whole dataset. In addition, the cities not covered under the scheme do not matter as they get dropped in the final estimation as we discuss later in the estimation strategy.

[22] The usual two-way fixed effects estimator fails due to two reasons. First, when the treatment effects are dynamic i.e., when the treatment effects can change over time. In such a case, previously treated units form a bad control group for units that are treated later. Second, the weights attached to the treatment effects depend on the number of periods that a unit is under treatment. See Roth *et al.* (2022) for a review.

model:

$$Y_{jkmt} = \delta_{km} + \delta_t + \delta_j + \sum_{\tau=-3,\tau\neq-1}^{\tau=3} \beta_\tau Route^{\tau}_{km} + \epsilon_{jkmt} \tag{3}$$

where Y is the number of applications received by job ad j in city k from candidates in city m with posting date t. The main variable of interest, $Route^{\tau}_{km}$ is an indicator variable that takes a value one for a city pair (k, m), τ periods after the route starts, and zero otherwise. Here, $\tau \in \{-3, 3\}$ and we normalize the coefficients relative to $\tau = -1$, i.e. the quarter before the treatment quarter. The quarter in which the route becomes operational corresponds to $\tau = 0$.[23] β_τ measures the average treatment effect on the number of applications in τ. The event study design allows for direct testing of common pre-trends assumption. Specifically, we test whether β_τ before the opening of routes is significantly different from zero.

We control for city pair level fixed-effects (δ_{km}), time fixed-effects (δ_t) and job level fixed-effects (δ_j). δ_{km} accounts for non-time-varying heterogeneity for city pair (k, m), while δ_j controls for any job-specific variation. Given these extensive controls, the staggered DID method proposed by Sun & Abraham (2021) that allows for adding covariates is best suited for our case. We cluster the standard errors at job and candidate city level (k, m) given that the treatment is at the level of city pair.

Figure 2 plots the coefficients from the event study estimates. Panel (a) gives the impact on jobs located in metro cities, while Panel (b) plots it for jobs in non-metro cities. By conducting this analysis separately for these city groups, we ensure that the control group consists of comparable job location cities, primarily in terms of city size. We find that when candidates in smaller cities are linked to metro cities via *UDAN* initiative, there is a significant increase in applications, with an increase of roughly 0.02 applications per job. Given the average application rate is 0.078 in metro cities, this constitutes a 25 percent increase in

[23] The leads and lags are defined in the following manner. $\tau = 0$ is the treatment quarter where the month when the route becomes operational is the mid-month for $\tau = 0$. $\tau = -1$ is the quarter before, $\tau = -2$ is two-three quarters before, and $\tau = -3$ is three quarters before the treatment quarter ($\tau = 0$), respectively. Similarly, $\tau = 1$ is the quarter after, $\tau = 2$ is two-three quarters after, $\tau = 3$ is beyond three quarters after $\tau = 0$. We create bins for the endpoints of the event window based on standard event-study applications (e.g. Schmidheiny & Siegloch (2019)).

application for jobs located in a distant metro. Concurrently, no positive pre-trends are visible prior to the commencement of these routes. However, in Panel (b), there is no significant increase in applications from non-metro to other non-metro cities, even if they get connected via the scheme.[24]

One concern with causal interpretation of these results could come from the endogeneity of route selection under the scheme. As our sample is limited to the subset of candidate locations and job locations that have received at least one route, this concern is largely mitigated. However, when city k gets connected to m_1 and not to m_2, there could be a concern that starting the route between (k, m_1) can be due to unobserved factors that also influence applications over time. For instance, we find that more distant city pairs, among sample of those at least 200 miles apart, have a lower likelighood to be connected through *UDAN*. Therefore, we control for trends in applications across cities over time due to their distance from each other. Figure 3 plots these estimates after controlling for time fixed-effects based on distance between city pairs. We find that the results remain consistent. In fact, the effects are more significant for jobs in metro locations. We note some positive impact of the scheme for jobs located in non-metro cities, but only beyond three quarter ($\tau = 2$) after the treatment.

To sum up, our findings underscore the importance of connectivity costs in influencing job applications to metropolitan cities. If a city gets an air route to a metro, it increases the likelihood of candidates applying to jobs in the linked metro. However, the connectivity costs do not seem to explain the observed distaste for jobs located in non-metro cities.

5.2 City-Specific Concentration of Job Roles

Another potential influence on candidate applications could be the concentration of specific job roles within individual cities. On average, metro cities have more job opportunities (the five metros account for 60 percent of job postings on the portal), which can incentivize

[24] We find no overall effect if we estimate the impact on all job locations, both metro and non-metro, in a single specification.

candidates to favor these locations due to the perceived option value associated with better match in future job search. For example, if finance-related jobs are predominantly located in Mumbai, candidates planning a career in this field may choose Mumbai considering better future opportunities. Such role-based concentration may also exist for non-metro location, thus reducing the distaste for distance for particular non-metro$\times$job role pairs.

We test this hypothesis and report the results in Table 3. Using the job role information from our dataset, we compute the proportion of jobs in a specific job role posted in a city during July–December 2020. For each job role, the sum of proportions across all cities adds up to one. We then use a specification based on Equation 2 to assess the impact of distance$\times$proportion on the likelihood of candidate application. Column (1) reports the baseline results. All coefficients related to the interaction terms on distance and proportion are positive and significant, showing a preference among candidates for cities with a higher proportion of job postings in a given role. With a 10 percentage point increase in the proportion of jobs in a specific role in a city, a candidate's likelihood of applying to a job over 500 miles away increases by 1.9 percentage points. There continues to be a large distaste for non-metro locations, however, some distaste towards non-metro locations could be offset if the location specializes in a particular job role.

In column (1), we account for unobserved heterogeneity at job and candidate levels, effectively eliminating any effect from time-invariant attributes for job and candidate cities. However, the static proportion variable for a specific job role-city pair may correlate with other city characteristics, thus failing to capture the causal effect of job role concentration within a city. To address this, we define the proportion of jobs posted in a city for a given job role at the candidate-level, based on the proportion of job advertisements visible to a specific candidate on the portal. This approach is based the idea that candidates may revise their perceptions about a job role-city pair based on the job ads they see on the portal on the day of their visit. Since the candidate-level proportion varies over time for a job role-city pair, our identification would come from comparing application probabilities based on the fluctuating proportion of ads in a job role-city pair over time.

We use candidate's choice set, as outlined in Section 3.1, i.e., job roles posted between $d - 10$ and $d + 1$, to calculate the candidate-level proportion. The results with this variable are reported in columns (2)–(5). Again, we find that an increase in concentration of a job role in a city increases the probability of application from candidates (column (2)). A candidate is 0.54 percentage points more likely to apply to a distant location (> 500 miles) when there is a 10 percentage point increase in proportion of jobs in given job role-city pair. We obtain similar results in column (3) when we saturate the specification by additionally incorporating job role×city×distance fixed-effects. These fixed-effects filter out any time-invariant characteristics for a given job role-city pair over a given distance.

Lastly, we examine whether job concentration has a varying impact on the application probability for metro vs. non-metro jobs. We split the sample by metro and non-metro job locations and report the results in the subsequent two columns (4) & (5). A high concentration of job roles increases the application probability for both metro and non-metro cities. The mediating effect of job concentration on application to non-metro locations seem more substantial. For distances exceeding 500 miles, a 10 percentage point increase in proportion of a job role in metro cities increases the application probability by 0.39 percentage points, roughly (0.0039/.053 =) 7.3 percent of the baseline decline in probability. A similar change in proportion in non-metro locations corresponds to a (0.025/0.15 =) 16.6 percent increase in application probability. This effect is on a larger baseline distaste of 15 percent.

Juxtaposing these findings against the long-distance connectivity scheme show that long-term job opportunities, as reflected in the concentration of job roles in a city, may be more meaningful in reducing candidate distaste for non-metro locations.

5.3 Other City-specific Attributes

In Table 4, we account for the differential effect of other factors that could potentially explain distaste for non-metro jobs. The job's location within the candidate's home state (column

(1)), in a bordering state (column (2)), or having direct language proximity (column (3)) do not seem to explain the higher distaste for more distant non-metro locations. The interaction of distance with these characteristics is significant in many cases. However, these factors don't adequately explain the higher distaste for distant non-metro jobs, as evidenced by the minimal reduction in the distaste for non-metro cities even after their effect is accounted for.

5.4 Employer Distaste for Distant Candidates

Our analyses thus far has looked at candidates' preferences in generating the observed distaste for distant jobs. However, such distaste could be an equilibrium result where employers also exhibit a similar distaste for candidates residing far away, thereby making candidates' choice to not apply to these jobs a rational choice. In this subsection, we directly examine if employers demonstrate any such distaste, and if there is any heterogeneity between employers located in metro vs. non-metro cities. We estimate the following equation:

$$Shortlist_{ijkt} = \gamma_0 + \sum_{k=1}^{g} \gamma_k DistGroup_k + \lambda_i \mathbb{X}_i + \lambda_j \mathbb{X}_j + \mu_i + \mu_j + \mu_t + \varepsilon_{ijkt} \tag{4}$$

where $Shortlist_{ijkt}$ is a binary variable that equals one if candidate i is shortlisted by the employer of the posted ad j located at a distance k in month-year t, and zero otherwise. Unlike in Equation 1, the choice set is well defined here since an employer shortlists over the given set of received applications. A negative coefficient γ_k suggests that firms are less likely to shortlist candidates located at a distance $DistGroup_k$ relative to candidates located in the same city as the posted job.

Incorporating μ_i and μ_j, the candidate and job fixed-effects, helps control for unobservable factors at the candidate and job level that can effect shortlisting. Therefore, our identification accounts for any differences in shortlisting that can arise from factors like candidate ability or job role description, among others. We also include vector of candidate- and job-level controls $\mathbb{X}_i$ and $\mathbb{X}_j$, but their coefficients would not be estimated when we include μ_i and μ_j

terms in the regression. Finally, μ_t controls for temporal changes in shortlisting. We weight the regressions by the inverse of total applications received for a posted job ad to give equal weight to all job ads in our regressions. The standard errors are clustered at the job level.

Table 7 reports these results. Initially, we only control for month-year fixed effects in shortlisting by including month-year fixed-effects (column (1)). We find that the coefficients on distance groups above 50 miles are negative and significant showing that employers are less likely to shortlist candidates located far away. However, once we introduce candidate and job fixed-effects, the decline in shortlisting probability is modest, standing at 4.1 percentage points, or by 14 percent, for candidates located more than 500 miles away (column (2)). No distaste is seen for candidates located upto 500 miles. Juxtaposing our findings for employer distaste with the candidate distaste reveals two key insights. First, candidate display a significant distaste for distance, which rises exponentially for jobs located beyond 100 miles. For employers, the distaste is non-existent until 500 miles and matters only when the candidate is located beyond this distance (column (4)). Second, the candidate distaste for distance is quantitatively large, reducing the probability of application to jobs beyond 100 miles to nearly zero. The employers distaste only reduces the probability of shortlisting candidates beyond 500 miles by 14 percent. Thus, the distaste for distance displayed by job seekers is likely driven by supply side preferences rather than an equilibrium outcome predicated on employer distaste towards distant candidates.[25]

Lastly, we test if employers also display heterogeneity in their distaste for distant candidates based on whether the job is located in metro or non-metro (columns (3)–(5)). All specifications show that employers hiring for jobs in metro locations are less likely to shortlist candidates located at least 500 miles away by 6-9 percentage points, around 26 percent. Conversely, the coefficients on the interaction terms with distance groups above 100 miles and a non-metro indicator variable are positive and significant. Quantitatively, employers are 11 percentage

[25]The observed employer distaste for candidates located over 500 miles away could partially be attributed to a hesitance to shortlist female candidates from distant locations. This may stem from an assumption that female candidates from far-off cities might be less likely to accept an offer if selected. Appendix Table B.8 reports these results.

point more likely to shortlist candidates located 500 miles away for non-metro jobs than metro jobs (column (4)). This effect nullifies the baseline distaste for candidates beyond 500 miles in case of jobs located in non-metro cities. These findings are robust to including candidate fixed-effects in column (5) and heterogeneous effects on shortlisting due to other job level attributes like education and experience requirements in column (6).[26] This result once again shows that candidate distaste for non-metro jobs is not reciprocated by a similar distaste on the employer side.

6 Conclusion

Our paper uses data from a large online job portal in India to study the candidate distaste for distant jobs. We find that job seekers may not have the same distaste for all distant cities – they are more likely to apply to distant metro cities than non-metro cities. We then test for the importance of various factors in explaining this heterogeneous distaste. To test the role played by connectivity costs, we exploit a policy change that increased subsidized air connectivity of smaller cities. Using a staggered differences-in-differences strategy, we find that a decrease in connectivity costs increases applications from smaller to larger metro cities but has no effect on applications towards jobs in smaller cities. These findings show that connectivity costs or infrastructure may not be the binding constraint for tighter labor markets in smaller cities. Instead, lower job concentration in smaller cities which reduces future job mobility can be an important factor contributing to the higher relative distaste for non-metros. We rule out cultural proximity as a potential mechanism and also find that a higher potential employer distaste for distant candidates does not drive the observed spatial patterns in candidate applications.

[26] We report the robustness of our results to including jobs with job roles as 'others' in Appendix Table B.9. Our previous results continue to hold.

References

Ades, Alberto F, & Glaeser, Edward L. 1995. Trade and circuses: explaining urban giants. *The Quarterly Journal of Economics*, **110**(1), 195–227.

Afridi, Farzana, Mahajan, Kanika, & Sangwan, Nikita. 2021. The gendered effects of climate change: Production shocks and labor response in agriculture.

Albouy, David, & Stuart, Bryan A. 2020. Urban population and amenities: The neoclassical model of location. *International economic review*, **61**(1), 127–158.

Albouy, David, Christensen, Peter, & Sarmiento-Barbieri, Ignacio. 2020. Unlocking amenities: Estimating public good complementarity. *Journal of Public Economics*, **182**, 104110.

Banerjee, Abhijit, Duflo, Esther, & Qian, Nancy. 2020. On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, **145**, 102442.

Banfi, Stefano, & Villena-Roldan, Benjamin. 2019. Do high-wage jobs attract more applicants? Directed search evidence from the online labor market. *Journal of Labor Economics*, **37**(3), 715–746.

Banfi, Stefano, Choi, Sekyu, & Villena-Roldán, Benjamin. 2019. Deconstructing job search behavior. *Available at SSRN 3323545*.

Baum-Snow, Nathaniel. 2007. Did highways cause suburbanization? *The quarterly journal of economics*, **122**(2), 775–805.

Blonigen, Bruce A, & Cristea, Anca D. 2015. Air service and urban growth: Evidence from a quasi-natural policy experiment. *Journal of Urban Economics*, **86**, 128–146.

Breidenbach, Philipp. 2020. Ready for take-off? The economic effects of regional airport expansions in Germany. *Regional Studies*, **54**(8), 1084–1097.

Brenčič, V. 2012. Wage posting: Evidence from job ads. *Canadian Journal of Economics*, **45(4)**, 1529–59.

Callaway, Brantly, & Sant'Anna, Pedro HC. 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics*, **225**(2), 200–230.

Carlino, Gerald A, & Saiz, Albert. 2019. Beautiful city: Leisure amenities and urban growth. *Journal of Regional Science*, **59**(3), 369–408.

Chaturvedi, Sugat, Mahajan, Kanika, & Siddique, Zahra. 2021. *Words matter: Gender, jobs and applicant behavior*. Tech. rept. IZA Discussion Paper.

Chauvin, Juan Pablo, Glaeser, Edward, Ma, Yueran, & Tobio, Kristina. 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, **98**, 17–49.

Clark, Terry Nichols, Lloyd, Richard, Wong, Kenneth K, & Jain, Pushpam. 2002. Amenities drive urban growth. *Journal of urban affairs*, **24**(5), 493–515.

Davis, Steven J, & de la Parra, Brenda Samaniego. 2017. Application flows. *Unpublished manuscript*.

Desmet, Klaus, Ghani, Ejaz, O'Connell, Stephen, & Rossi-Hansberg, Esteban. 2015. The spatial development of India. *Journal of Regional Science*, **55**(1), 10–30.

Donaldson, Dave. 2018. Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, **108**(4-5), 899–934.

Duranton, Gilles, & Turner, Matthew A. 2012. Urban growth and transportation. *Review of Economic Studies*, **79**(4), 1407–1440.

Eeckhout, Jan, Pinheiro, Roberto, & Schmidheiny, Kurt. 2014. Spatial sorting. *Journal of Political Economy*, **122**(3), 554–620.

Eriksson, Tor, & Kristensen, Nicolai. 2014. Wages or fringes? Some evidence on trade-offs and sorting. *Journal of Labor Economics*, **32**(4), 899–928.

Faberman, R Jason, & Kudlyak, Marianna. 2019. The intensity of job search and search duration. *American Economic Journal: Macroeconomics*, **11**(3), 327–57.

Fluchtmann, Jonas, Glenny, Anita, Harmon, Nikolaj A, & Maibom, Jonas. 2021. The Gender Application Gap: Do men and women apply for the same jobs?

Fujiwara, Thomas, Morales, Eduardo, & Porcher, Charly. 2022. A Revealed-Preference Approach to Measuring Information Frictions in Migration Decisions. *Working Paper*.

Gibbons, Stephen, & Wu, Wenjie. 2020. Airports, access and local economic performance: Evidence from China. *Journal of Economic Geography*, **20**(4), 903–937.

Glaeser, Edward L, & Mare, David C. 2001. Cities and skills. *Journal of labor economics*, **19**(2), 316–342.

Glaeser, Edward L, & Resseger, Matthew G. 2010. The complementarity between cities and skills. *Journal of Regional Science*, **50**(1), 221–244.

Gollin, Douglas, Kirchberger, Martina, & Lagakos, David. 2021. Do urban wage premia reflect lower amenities? Evidence from Africa. *Journal of Urban Economics*, **121**, 103301.

Goodman-Bacon, Andrew. 2021. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, **225**(2), 254–277.

He, Haoran, Neumark, David, & Weng, Qian. 2021. Do workers value flexible jobs? A field experiment. *Journal of Labor Economics*, **39**(3), 709–738.

Henderson, J Vernon. 1986. Urbanization in a developing country: City size and population composition. *Journal of Development Economics*, **22**(2), 269–293.

Hensvik, Lena, Le Barbanchon, Thomas, & Rathelot, Roland. 2021. Job search during the COVID-19 crisis. *Journal of Public Economics*, **194**, 104349.

Kudlyak, Marianna, Lkhagvasuren, Damba, & Sysuyev, Roman. 2012. Systematic Job Search: New Evidence from Individual Job Application Data, Working Paper 12-03R.

Kuhn, Peter, & Shen, Kailing. 2013. Gender discrimination in job ads: Evidence from china. *The Quarterly Journal of Economics*, **128**(1), 287–336.

Kuhn, Peter, Shen, Kailing, & Zhang, Shuo. 2020. Gender-targeted job ads in the recruitment process: Facts from a Chinese job board. *Journal of Development Economics*, 102531.

Kuhn, Peter J, & Shen, Kailing. 2021. *What Happens When Employers Can No Longer Discriminate in Job Ads?* Tech. rept. National Bureau of Economic Research.

Le Barbanchon, Thomas, Rathelot, Roland, & Roulet, Alexandra. 2021. Gender differences in job search: Trading off commute against wage. *The Quarterly Journal of Economics*, **136**(1), 381–426.

Lin, Yatang. 2017. Travel costs and urban specialization patterns: Evidence from Chinaâs high speed railway system. *Journal of Urban Economics*, **98**, 98–123.

Ma, Ke Amy, & de la Parra, Brenda Samaniego. 2021. *Labor Market Tightness, Recruitment and Search Behavior*. Tech. rept. Working Paper.

Marinescu, Ioana, & Rathelot, Roland. 2018. Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics*, **10**(3), 42–70.

Marinescu, Ioana, & Wolthoff, Ronald. 2020. Opening the black box of the matching function: The power of words. *Journal of Labor Economics*, **38**(2), 535–568.

Mas, Alexandre, & Pallais, Amanda. 2017. Valuing alternative work arrangements. *American Economic Review*, **107**(12), 3722–59.

McGraw, Marquise J. 2020. The role of airports in city employment growth, 1950–2010. *Journal of Urban Economics*, **116**, 103240.

Michaels, Guy. 2008. The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, **90**(4), 683–701.

Michelacci, C., & Suarez, J. 2006. Incomplete Wage Posting. *Journal of Political Economy*, **114(6)**, 1098â123.

Phillips, David C. 2020. Do low-wage employers discriminate against applicants with long commutes? Evidence from a correspondence experiment. *Journal of Human Resources*, **55**(3), 864–901.

Porcher, Charly. 2020. Migration with Costly Information.

Rickne, Johanna, & Folke, Olle. 2020. Sexual Harassment and Gender Inequality in the Labor Market.

Roth, Jonathan, Sant'Anna, Pedro HC, Bilinski, Alyssa, & Poe, John. 2022. What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *arXiv preprint arXiv:2201.01194*.

Schmidheiny, Kurt, & Siegloch, Sebastian. 2019. On event study designs and distributed-lag models: Equivalence, generalization and practical implications.

Sheard, Nicholas. 2014. Airports and urban sectoral employment. *Journal of Urban Economics*, **80**, 133–152.

Sun, Liyang, & Abraham, Sarah. 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, **225**(2), 175–199.

Van Nierop, Erjen, Bronnenberg, Bart, Paap, Richard, Wedel, Michel, & Franses, Philip Hans. 2010. Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research*, **47**(1), 63–74.

Wiswall, Matthew, & Zafar, Basit. 2018. Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, **133**(1), 457–507.

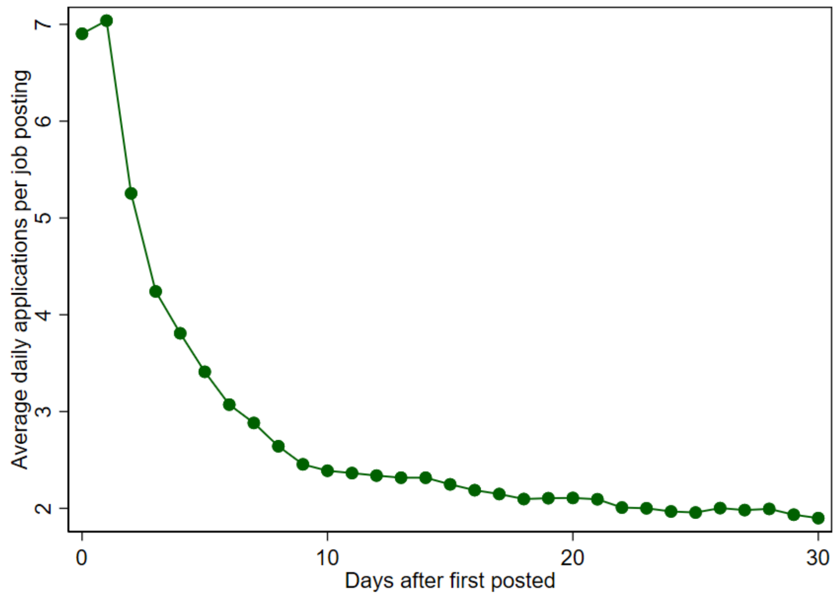
Wozniak, Abigail. 2010. Are college graduates more responsive to distant labor market opportunities? *Journal of Human Resources*, **45**(4), 944–970.

Yakubenko, Slava. 2020. Giants and midgets: The effect of public goods provision on urban population concentration. *Cities*, **107**, 102872.

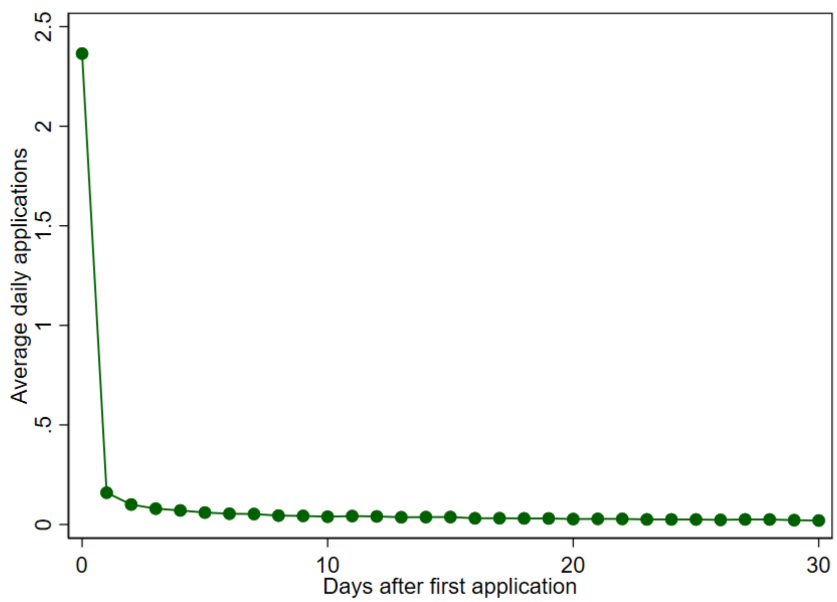
Zhang, Hui, & Xie, Tingting. 2023. A key to urban economic growth or an unnecessary burden? Opening airports in small and medium-sized cities. *Cities*, **133**, 104105.

Figure 1: Activity on Posted Jobs

(a) Average Applications from Date of Job Posting



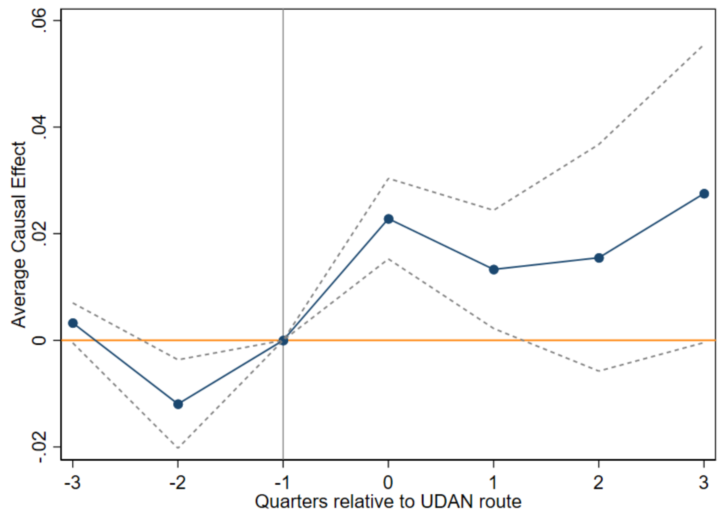
(b) Average Daily Applications by Candidates after First Application



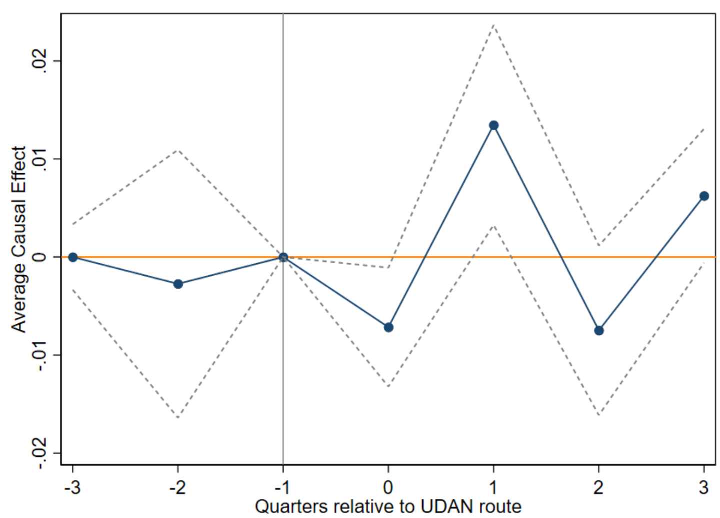
Notes: Panel (a) shows the daily average of applications received on a job since it's posting date for the period July-December 2020. Jobs receive maximum applications immediately after the job posting and it falls almost flat beyond the 10 days period. Here, Day0 is defined as job posting date. Same day applications would be clubbed together under Day0. Similarly applications received after that will be clubbed on the basis of gap days between the two dates (application vs job posting). Chart shows mean number of applications by gap days. Panel (b) shows the daily average applications by candidates for the period July-December 2020. When a candidate applies for the first time in the considered time period has been defined as Day0. All applications beyond Day0 are clubbed together on the basis of gap days.

Figure 2: Effect of Increased Connectivity under Udan Scheme on Applications

(a) Metro jobs



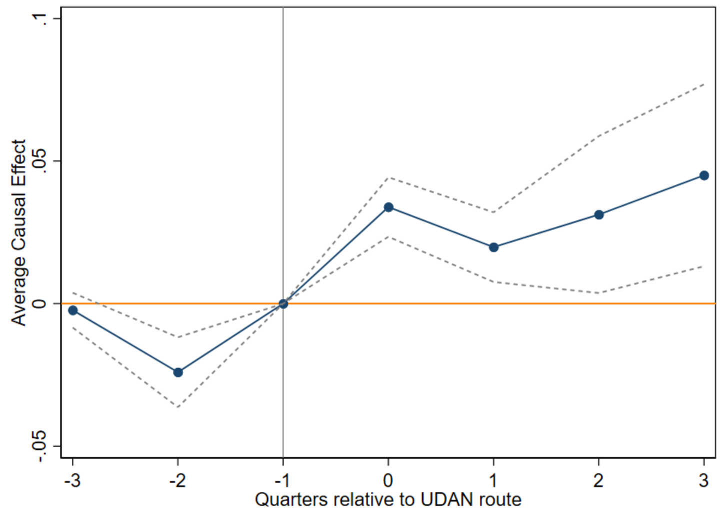
(b) Non-metro jobs



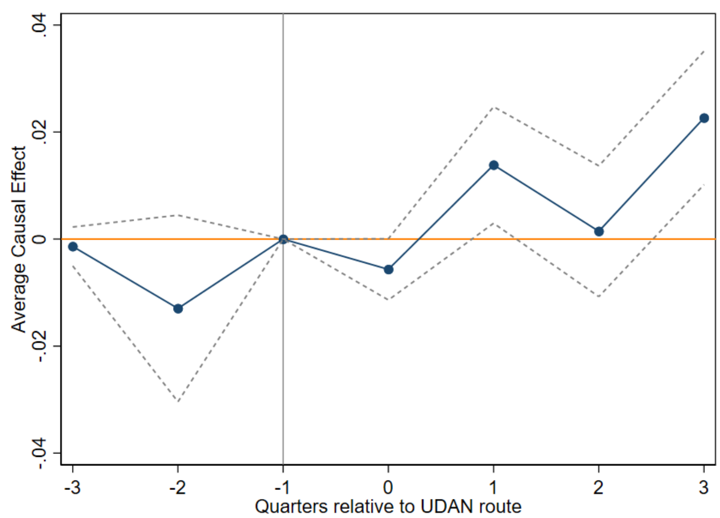
Notes: We plot the impact of air routes becoming operational between a city pair on applications received for the posted job. The outcome of interest is the number of applications received by a job in city k from candidates in city m. Event-study plots using the Sun & Abraham (2021) estimator. The solid line represents the average annual treatment effects, and the dashed lines denote the 90% confidence intervals. The treatment effects are with respect to the quarter before the route came into force (dashed vertical line). Both specifications include city pair, posting date and job ad fixed effects. Standard errors are clustered at city pair level.

Figure 3: Effect of Increased Connectivity under Udan Scheme on Applications (Robustness: Distance time trends)

(a) Metro jobs



(b) Non-metro jobs



Notes: We plot the impact of air routes becoming operational between a city pair on applications received for the posted job. The outcome of interest is the number of applications received by a job in city k from candidates in city m. Event-study plots using the Sun & Abraham (2021) estimator. The solid line represents the average annual treatment effects, and the dashed lines denote the 90% confidence intervals. The treatment effects are with respect to the quarter before the route came into force (dashed vertical line). Both specifications include city pair, posting date and job ad fixed effects. Additionally, we also control for interaction of distance between a city pair and month-year fixed effects. Standard errors are clustered at city pair level.

Table 1: Effect of Distance on Application by candidates

	(1)	(2)	(3)	(4)
1-50	–0.014***	–0.015***	–0.017***	–0.015***
	(0.002)	(0.001)	(0.002)	(0.001)
50-100	–0.048***	–0.048***	–0.053***	–0.051***
	(0.001)	(0.000)	(0.000)	(0.000)
100-500	–0.058***	–0.056***	–0.062***	–0.059***
	(0.000)	(0.000)	(0.000)	(0.000)
> 500	–0.061***	–0.059***	–0.065***	–0.062***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.065***	0.069***	0.073***	0.065***
	(0.000)	(0.002)	(0.001)	(0.000)
N	166555125	164769978	166487421	166555050
Mean Y	0.010	0.010	0.010	0.010
Mean Y (Same city)	0.065	0.065	0.065	0.065
Controls				
Job controls			✓	
Candidate controls		✓		
Job FE		✓		✓
Candidate FE			✓	✓
Month-year FE	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table 2: Effect of Distance on Application by candidates: Heterogeneity by metro vs non-metro job location

Sample:	All jobs	Jobs with posted wages		
	(1)	(2)	(3)	(4)
1-50	–0.013***	–0.022***	–0.036***	–0.023***
	(0.001)	(0.002)	(0.003)	(0.003)
50-100	–0.032***	–0.053***	–0.063***	–0.056***
	(0.000)	(0.001)	(0.001)	(0.001)
100-500	–0.041***	–0.066***	–0.076***	–0.069***
	(0.000)	(0.000)	(0.001)	(0.001)
> 500	–0.046***	–0.075***	–0.085***	–0.078***
	(0.000)	(0.000)	(0.001)	(0.001)
1-50 x $\mathbb{1}$(Non-metro)	–0.045***	–0.111***	–0.109***	–0.155***
	(0.003)	(0.004)	(0.004)	(0.006)
50-100 x $\mathbb{1}$(Non-metro)	–0.073***	–0.147***	–0.146***	–0.181***
	(0.001)	(0.002)	(0.002)	(0.003)
100-500 x $\mathbb{1}$(Non-metro)	–0.073***	–0.152***	–0.152***	–0.186***
	(0.001)	(0.002)	(0.002)	(0.002)
> 500 x $\mathbb{1}$(Non-metro)	–0.074***	–0.150***	–0.150***	–0.184***
	(0.001)	(0.002)	(0.002)	(0.002)
1-50 x $\mathbb{1}$(wage> med)			0.026***	0.002
			(0.004)	(0.004)
50-100 x $\mathbb{1}$(wage> med)			0.020***	0.006***
			(0.001)	(0.001)
100-500 x $\mathbb{1}$(wage> med)			0.020***	0.006***
			(0.001)	(0.001)
> 500 x $\mathbb{1}$(wage> med)			0.019***	0.006***
			(0.001)	(0.001)
1-50 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.098***
				(0.008)
50-100 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.071***
				(0.003)
100-500 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.071***
				(0.003)
>500 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.069***
				(0.003)
Constant	0.078***	0.157***	0.146***	0.148***
	(0.000)	(0.001)	(0.001)	(0.001)
N	166555050	62575521	62575521	62575521
Mean Y	0.010	0.018	0.018	0.018
Mean Y (Same city)	0.065	0.108	0.108	0.108
Controls				
Job FE	✓	✓	✓	✓
Candidate FE	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table 3: Effect of Job Role Concentration on Application by Candidates

Proportion:	Aggregate	Candidate			
				Metro	Non-metro
	(1)	(2)	(3)	(4)	(5)
1-50	–0.044***	–0.032***		–0.015***	–0.119***
	(0.003)	(0.003)		(0.002)	(0.007)
50-100	–0.066***	–0.046***		–0.048***	–0.143***
	(0.001)	(0.001)		(0.001)	(0.001)
100-500	–0.078***	–0.055***		–0.051***	–0.155***
	(0.001)	(0.000)		(0.000)	(0.001)
> 500	–0.079***	–0.057***		–0.053***	–0.159***
	(0.001)	(0.000)		(0.000)	(0.001)
1-50 × Proportion	0.170***	0.084***	–0.029**	0.017***	1.004***
	(0.009)	(0.013)	(0.015)	(0.006)	(0.344)
50-100 × Proportion	0.187***	0.067***	0.018***	0.069***	0.279***
	(0.005)	(0.005)	(0.006)	(0.004)	(0.013)
100-500 × Proportion	0.213***	0.073***	0.034***	0.052***	0.317***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.006)
>500 × Proportion	0.189***	0.054***	0.035***	0.039***	0.254***
	(0.003)	(0.002)	(0.001)	(0.002)	(0.006)
1-50 × Non-metro	–0.032***	–0.033***			
	(0.004)	(0.005)			
50-100 × Non-metro	–0.056***	–0.067***			
	(0.001)	(0.001)			
100-500 × Non-metro	–0.056***	–0.067***			
	(0.001)	(0.001)			
>500 × Non-metro	–0.059***	–0.070***			
	(0.001)	(0.001)			
Constant	0.081***	0.079***	–0.696	0.049***	0.142***
	(0.000)	(0.000)	(77249.518)	(0.000)	(0.001)
N	166555050	166555050	166554401	102007876	64546823
Job FE	✓	✓	✓	✓	✓
Candidate FE	✓	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓	✓
Job role, city × Dist. groups			✓		

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table 4: Effect of Distance on Application by candidates (Metro vs Non-metro): Other factors

	Proportion	Within state	Border state	Language proximity
1-50	–0.020***	–0.029***	0.016	–0.021***
	(0.005)	(0.002)	(0.013)	(0.003)
50-100	–0.059***	–0.037***	–0.028***	–0.043***
	(0.001)	(0.001)	(0.002)	(0.001)
100-500	–0.067***	–0.041***	–0.045***	–0.041***
	(0.001)	(0.000)	(0.000)	(0.000)
> 500	–0.073***	–0.046***	–0.046***	–0.046***
	(0.001)	(0.000)	(0.000)	(0.000)
1-50 x $\mathbb{1}$(Non-metro)	–0.047***	–0.050***	–0.046***	–0.045***
	(0.005)	(0.003)	(0.003)	(0.003)
50-100 x $\mathbb{1}$(Non-metro)	–0.054***	–0.074***	–0.073***	–0.074***
	(0.001)	(0.001)	(0.001)	(0.001)
100-500 x $\mathbb{1}$(Non-metro)	–0.056***	–0.073***	–0.073***	–0.073***
	(0.001)	(0.001)	(0.001)	(0.001)
> 500 x $\mathbb{1}$(Non-metro)	–0.056***	–0.074***	–0.074***	–0.074***
	(0.001)	(0.001)	(0.001)	(0.001)
1-50 x $\mathbb{1}$(Characterstic)	0.016***	0.033***	–0.030**	0.012***
	(0.005)	(0.003)	(0.013)	(0.004)
50-100 x $\mathbb{1}$(Characterstic)	0.042***	0.007***	–0.004	0.024***
	(0.001)	(0.001)	(0.002)	(0.002)
100-500 x $\mathbb{1}$(Characterstic)	0.044***	0.009***	0.005***	0.003***
	(0.001)	(0.000)	(0.000)	(0.000)
>500 x $\mathbb{1}$(Characterstic)	0.044***	–0.003	0.001***	0.003
	(0.001)	(0.005)	(0.000)	(0.000)
Constant	0.081***	0.078***	0.078***	0.078***
	(0.000)	(0.000)	(0.000)	(0.000)
N	166555050	166555050	166555050	166555050
Candidate FE	✓	✓	✓	✓
Job FE	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Interaction controls include distance interacted for graduates above and experience of more than one year for candidates. Interaction controls include proportion of job role, same state, bordering state for candidate, language proximity × distance. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with wage information available location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table 5: Effect of Distance on Application by candidates: Robustness

	Within state	Border State	Language Proximity	Social Connectivity
1-50	–0.029***	0.006***	–0.021***	0.004*
	(0.002)	(0.002)	(0.003)	(0.002)
50-100	–0.037***	–0.030***	–0.044***	–0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
100-500	–0.041***	–0.041***	–0.041***	–0.013***
	(0.000)	(0.000)	(0.000)	(0.001)
$>$ 500	–0.046***	–0.046***	–0.046***	–0.018***
	(0.000)	(0.000)	(0.000)	(0.001)
1-50 x $\mathbb{1}$(Non-metro)	–0.050***	–0.053***	–0.045***	–0.018***
	(0.003)	(0.003)	(0.003)	(0.003)
50-100 x $\mathbb{1}$(Non-metro)	–0.074***	–0.074***	–0.074***	–0.032***
	(0.001)	(0.001)	(0.001)	(0.001)
100-500 x $\mathbb{1}$(Non-metro)	–0.073***	–0.074***	–0.073***	–0.034***
	(0.001)	(0.001)	(0.001)	(0.001)
$>$ 500 x $\mathbb{1}$(Non-metro)	–0.074***	–0.074***	–0.074***	–0.034***
	(0.001)	(0.001)	(0.001)	(0.001)
1-50 x $\mathbb{1}$(Characterstic)	0.033***	–0.038***	0.013***	0.496
	(0.003)	(0.002)	(0.004)	(0.340)
50-100 x $\mathbb{1}$(Characterstic)	0.007***	–0.009***	0.024***	1.383***
	(0.001)	(0.001)	(0.002)	(0.447)
100-500 x $\mathbb{1}$(Characterstic)	0.009***	0.001***	0.003***	2.321***
	(0.000)	(0.000)	(0.000)	(0.322)
$>$500 x $\mathbb{1}$(Characterstic)	–0.003	0.001***	0.003***	13.717***
	(0.005)	(0.000)	(0.001)	(0.556)
Constant	0.078***	0.078***	0.078***	0.031***
	(0.000)	(0.000)	(0.000)	(0.001)
N	166555050	166555050	166555050	165960548
Candidate FE	✓	✓	✓	✓
Job FE	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Interaction controls include distance interacted for graduates above and experience of more than one year for candidates. Interaction controls include proportion of job role, same state, bordering state for candidate, language proximity × distance. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with wage information available location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table 6: Effect of Distance on Application by candidates

Candidate Proportion	(1)	Metro (2)	Non-metro (3)	Metro (4)	Non-metro (5)
1-50	–0.594	–0.015***	–0.119***	0.041	3.899
	(231852.625)	(0.002)	(0.007)	(15841.073)	(419789.450)
50-100	–0.271	–0.048***	–0.143***	–0.677	2.628
	(122757.285)	(0.001)	(0.001)	(129798.966)	(252245.035)
100-500	0.581	–0.051***	–0.155***	–0.179	–1.129
	(87292.091)	(0.000)	(0.001)	(26078.249)	(68539.557)
> 500	0.149	–0.053***	–0.159***	0.332	–0.099
	(65596.003)	(0.000)	(0.001)	(49116.159)	(64534.034)
1-50 x Proportion	–0.029**	0.017***	1.004***	–0.040***	0.055
	(0.015)	(0.006)	(0.344)	(0.011)	(0.390)
50-100 x Proportion	0.018***	0.069***	0.279***	0.002	0.058***
	(0.006)	(0.004)	(0.013)	(0.006)	(0.011)
100-500 x Proportion	0.034***	0.052***	0.317***	0.030***	0.067***
	(0.002)	(0.002)	(0.006)	(0.002)	(0.005)
>500 x Proportion	0.035***	0.039***	0.254***	0.031***	0.069***
	(0.001)	(0.002)	(0.006)	(0.002)	(0.004)
Constant	–0.465	0.049***	0.142***	–0.144	0.396
	(77249.520)	(0.000)	(0.001)	(25732.395)	(23207.071)
N	166554401	102007876	64546823	102007865	64546179
Job FE	✓	✓	✓	✓	✓
Candidate FE	✓	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓	✓
Job role, city × Dist. groups	✓			✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table 7: Effect of Distance on Shortlisting by Employer: Heterogeneity

	(1)	(2)	(3)	(4)	(5)
1-50	0.062**	0.010	0.080	0.149	0.147
	(0.027)	(0.067)	(0.048)	(0.116)	(0.119)
50-100	–0.052**	–0.035	0.022	–0.011	0.015
	(0.022)	(0.066)	(0.034)	(0.083)	(0.082)
100-500	–0.151***	–0.001	–0.020	–0.059**	–0.049
	(0.013)	(0.022)	(0.016)	(0.028)	(0.032)
> 500	–0.195***	–0.049**	–0.066***	–0.099***	–0.081***
	(0.014)	(0.020)	(0.015)	(0.025)	(0.029)
1-50 x $\mathbb{1}$(Non-metro)			–0.128**	–0.225	–0.230
			(0.063)	(0.157)	(0.155)
50-100 x $\mathbb{1}$(Non-metro)			0.007	–0.024	0.023
			(0.046)	(0.161)	(0.148)
100-500 x $\mathbb{1}$(Non-metro)			0.070***	0.147***	0.139***
			(0.025)	(0.047)	(0.047)
> 500 x $\mathbb{1}$(Non-metro)			0.112***	0.166***	0.168***
			(0.026)	(0.045)	(0.045)
Constant	0.359***	0.324***	0.210	0.332***	0.329***
	(0.010)	(0.008)	(0.135)	(0.010)	(0.010)
N	56410	28158	55738	28158	28158
Mean Y	0.295	0.295	0.295	0.295	0.295
Mean Y (Same city, base)	0.357	0.357	0.393	0.393	0.417
Controls					
Candidate controls			✓		
Job Edu, Exp × Distance					✓
Job FE		✓	✓	✓	✓
Candidate FE		✓		✓	✓
Month-year FE	✓	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate was shortlisted by an employer for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Candidate controls include gender, education, age and age square and city location of the candidate. Regressions weighted by inverse of total applicants to a posted job. Each column reports the effective number of observations after incorporating the included fixed effects. Robust standard errors clustered at the job level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any shortlisting was done by the employers during July 2020 - December 2020.

Table 8: Applications: Metro vs Non-Metro

	(1)	(2)	(3)	(4)	(5)
Non-metro	–54.473***	–55.470**	–39.870*	–8.774***	–9.845***
	(20.259)	(21.869)	(22.928)	(3.228)	(3.574)
log wage		–3.992			7.799**
		(14.964)			(3.299)
Constant	33.968***	82.494	20.421***	29.332***	–63.148
	(4.518)	(185.876)	(2.990)	(1.525)	(40.087)
N	881	879	1034	19138	16670
Mean Y	27.2	26.2	15.5	26.6	29.6
Controls					
Job Controls	✓	✓	✓	✓	✓
Job Role FE	✓	✓	✓	✓	✓
Job Title FE			✓	✓	✓
Month-year FE	✓	✓	✓	✓	✓

Notes: The dependent variable is applications received per vacancy in a job ad. Column (1) and (2) include jobs where any shortlisting was undertaken exluding those with job roles as 'others'. Column (3) includes jobs where any shortlisting was undertaken irrespective of job role. Column (4) and (5) include all jobs for all job roles where any applications was received and which were not cross-posted. Job controls include controls for education, experience, priority status of jobs and job location state. All specifications control for month and year of job posting. Each column reports the effective number of observations after incorporating the included fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city, during July 2020 - December 2020.

A Appendix: Data Details

First, we observe information on the posted jobs: date of posting, job type (full time vs part-time/internships), job role, job title, education requirements of the job ad, minimum and maximum required experience, minimum and maximum offered wage, city of location of the job ad, firm that posted the job ad, and other characteristics like number of vacancies. We only consider job ads which offer full-time paid opportunities. There were 51,549 full-time job ads posted on the platform during July 2020-December 2020. Further, we restrict our analyses to jobs for which only one city location is posted since only these jobs are feasible for our analyses. These constitute 90% of the posted jobs spanning across 449 cities and reduces our sample of jobs to 46,119. We then keep the job ads to which at least one candidate in our data applied during July-December 2020, further reducing the job set to 44,586 job ads. Dropping jobs with the job role as 'Others' gives us 37,770 job ads and we obtain a final set of 37,045 job ads across 419 cities after removing those which have no application within the first ten days of posting. We elaborate the importance of this selection in creating the choice set for our analyses later. This data include all information that is visible to the candidates on the platform.

Table A.1 shows the descriptive statistics for these jobs ads. Approximately 40% of the jobs require minimum education as schooling, diploma is required in 10% jobs while graduate education is required in 42% jobs. Postgraduate degree or higher is required in only 4.5% jobs. In terms of experience requirements, these jobs largely cater to young job seekers who are inexperienced, with 50% and 33% jobs posting a requirement of 0-1 and 1-2 years of experience. Only 12% jobs need an experience of more than five years. In terms of location characteristics, we divide our set of cities into metro (Delhi, Mumbai, Chennai, Kolkata, and Bengaluru), and non-metro (all other cities than the five mentioned before). The metro cities have been the main center of economic activity and population hubs within their regions. For instance, Delhi and Mumbai are the largest cities in north and west India, respectively. The five metro cities account for 60% jobs in our dataset. Appendix Figure A.1 shows the distribution of cities across the country for which a vacancy is posted. On average each job ad receives 45 applications.

Around 56% of the jobs ads i.e., 20,784 post minimum and maximum offered wages. This proportion is much larger than other existing studies which use posted wages. For instance, wages are advertised in just 13.4% in Banfi *et al.* (2019) and Banfi & Villena-Roldan (2019), 16.4% of job ads in Kuhn & Shen (2013), 20% of job ads in Marinescu & Wolthoff (2020) for the U.S. and 24.8% of job ads in Brenčič (2012). In line with the existing studies, we also find that wages are less likely to be posted for ads having higher skill requirements like higher education and experience, thus, allowing for possibility of wage bargaining and negotiation in higher skill jobs (Brenčič, 2012; Banfi & Villena-Roldan, 2019; Michelacci & Suarez, 2006). Hence, while we recognize that our sample of job ads with wage information is a selected sample of largely low skilled jobs, it is much larger than the existing studies to allow one to gauge the effect of posted wages on distance for distaste. The posted annual wage on average in a job ad that specifies a wage is Rs 2,40,000.

Second, we observe candidate characteristics, like age, gender, education, experience and most importantly all the job applications made by them to the above posted ads. We drop candidates with incomplete profiles, specifically the ones who do not fill up their gender information.[1] Our final dataset contains information on 0.68 million applications made by 2,26,464 candidates located across 679 cities. Table A.2 shows the descriptive statistics for the applicants. A large proportion of candidates, approximately 68%, are graduates while another 14% are postgraduates. Thus, on average the candidates on the portal have high education. In terms of experience, 73% of the candidates have either none or less than one year of experience while another 8% have 1-2 years of experience, again showing that young job seekers (average age is 24 years) are the primary client base for this portal. Around 44% of the candidates reside in metro locations. On average candidates apply to 4 full-time, single

[1]We restrict our analyses to only those who have gender information available because candidates who do not display gender, in general, do not enter other characteristics like education and date of birth as well and also make much fewer applications ($\approx 1/10$ of those made by candidates who reveal gender) on the portal. Thus, these seem to be non-serious candidates.

city jobs with specified job roles in our data during this time period. There are differences in candidate characteristics across gender. While female job seekers constitute 35% of the seekers, they are on average more educated and younger with lower job experience but apply apply to similar number of jobs as male job seekers.

Third, we can see which candidates are shortlisted by the firms for a given posted job for a subset of job ads for which employers purchase this service from the portal. For this period we observe shortlisting for 1470 job ads with all job roles and 888 job ads after dropping job ads with job role as 'Others'. To be consistent with our analyses on the candidate side we use the 888 job ads for our main analyses but extend our analyses later to all jobs. Table A.3 shows the summary stats for these jobs. It can be seen that these jobs require lower education and experience viz all jobs posted on the portal during this period. Around 83% of these jobs are in metro location, almost 100% have a specified wage and on average have 63 applicants per job ad. This shows that employers hiring for low skill jobs in metro locations - the ones who receive more applications - are also more likely to subscribe to this service provided by the job portal. This is consistent with those employers subscribing to the service who on an average are likely to get more applications and thus, have a higher marginal benefit from paying for this service.

A.1 Distaste: Role of Posted Wages and Other Characteristics

As one of the first factors, we consider the role played by posted wages in influencing candidate distaste. experience a disutility associated with relocation to a distant location, higher wages could potentially serve as a compensating factor. For this examination, we focus on the subset of jobs featuring posted wages to test if higher posted wages reduce the distaste for distance exhibited by candidates. Around 56% of the job ads report wages in our data and we construct the potential choice set over these jobs for each candidate. This yeilds around 62 million candidate×job observations with an average application rate of 1.7%.[2]

[2]Jobs with posted wages are likely to see a higher application, even after controlling for other job-level characteristics and candidate fixed-effects. We do not report these results in the main paper, but they are

Table B.1 presents the findings pertaining to the impact of wages on distaste. Column (1) reports the direct effect of wages on application probability after controlling for the distance effect. Here, we control for candidate fixed-effects, job requirements like education, experience and location city, as well as monthly trends. In this specification, we are estimating the direct effect of posted wage on application probability, thus job fixed-effects cannot be included. The coefficient on log wage is positive and significant, with application probability increasing by 0.6 percentage point with 1 percent increase in posted wage. This translates to a 35 percent increase in application probability for every 1 percent increase in posted wage. Thus, among the job ads that post wages, the ones that post higher wages are more likely to receive an application. The magnitude is similar to Marinescu & Wolthoff (2020).

The second column introduces an interaction term between log wage and distance group to investigate if higher wages could offset distaste for distance. We include both candidate and job fixed-effects in this specification to control for any effect on application rate due to job and candidate characteristics. The interaction coefficients on the distance groups and log wage are positive and significant, suggesting that given the distance, candidates have a lower distaste for the job with higher posted wages. In columns (3)–(4) instead of controlling for absolute wages we include an indicator variable which takes a value of one if the posted wage in a job ad is more than the median wage keeping the specification otherwise analogous to columns (1)–(2). The results in column (3) mirror those in column (1) and we find that jobs with above median wage witness higher application rate. In column (4), we interact the above median wage dummy with distance groups and find that the interaction coefficients in each case are positive and significant. In terms of magnitude, we find that jobs more than 100 miles away, with above median posted wages witness a reduction in distaste by 18 percent (0.02/0.11). These findings remain robust even upon including interaction terms of distance groups with other candidate-level characteristics in column (5).

We further examine the heterogeneity in distaste by gender, education and experience of

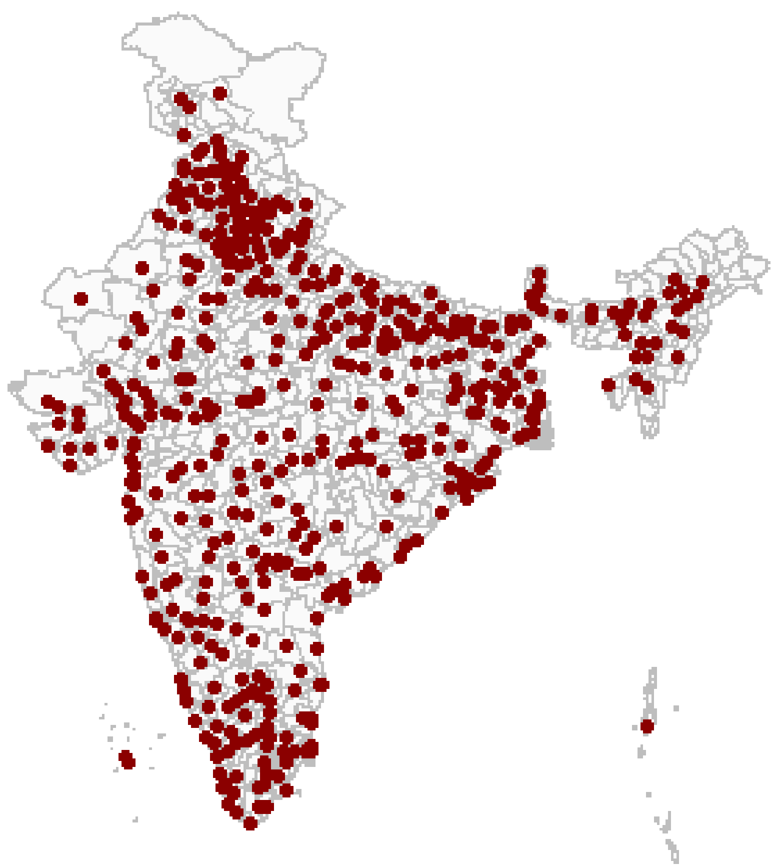
available on request.

the candidate (Appendix Table B.2). Column (1) illustrates the difference between male and female candidates, where we interact a binary variable *Female*, that takes a value of one for females and zero otherwise, with the distance groups. Female candidates exhibit a higher distaste for distance on average, as denoted by the negative and significant coefficients of the interaction terms. Given that female candidates are more educated than male candidates on the platform, we present estimates in column (2) for gender, after accommodating the differential effect of candidate education on preference for distant jobs. Here, the binary variable *Graduate* takes a value of one for candidates with a graduate degree or above, and zero otherwise. The interaction coefficients show that candidates with higher education demonstrate a lower distaste for distance. Lastly, column (3) incorporates an indicator variable, *Experienced*, that takes a value of one for candidates having at least six months of work experience, and zero otherwise. The estimate suggests that candidates with some experience have a higher distaste by approximately 17 percent.[3]

To summarize, these results reveal that while posted wages and other candidate attributes indeed influence application probability and the distaste towards distant jobs, they do not quantitatively account for the candidate's evident disinclination for jobs that are located far away. Despite including additional terms based on these characteristics in the regressions, the baseline coefficients on distance terms do not exhibit any significant change in their magnitude. This leads us to investigate the next key attribute, city type, in generating the distaste.

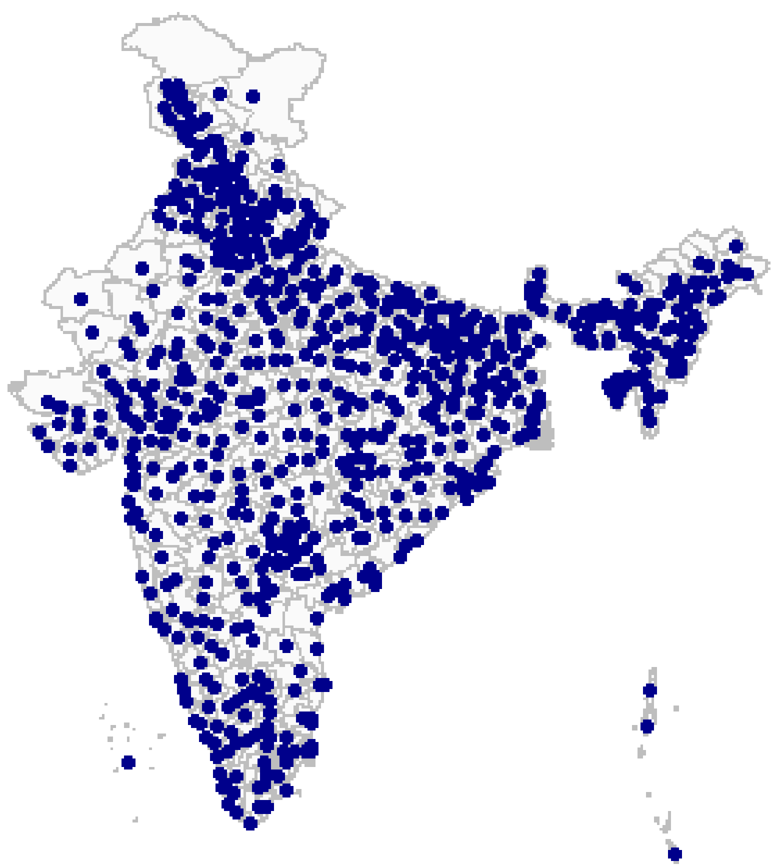
[3]Our findings on gender align with the studies that find lower average mobility among females than males in the Indian context and elsewhere (Le Barbanchon *et al.*, 2021; Afridi *et al.*, 2021). In contrast, Banfi *et al.* (2019) do not find any differential behavior in application rate by gender to distant jobs for Chile. The higher mobility of more educated workers has been documented for the US (Marinescu & Rathelot, 2018; Wozniak, 2010). For this job portal, experienced candidates are likely to be those who are currently employed. These results could reflect that those who search on the job may have a higher distaste for distant jobs since they already have employment in their current location.

Figure A.1: Location of jobs



Notes: The map plots the locations for which jobs are posted in our data for the period July - December 2020.

Figure A.2: Location of candidates



Notes: The map plots the locations for the candidates who apply to the posted jobs over the period July - December 2020.

Table A.1: Descriptive statistics: Job ads

	Mean	**SD**	**N**
Education:			
Secondary education or less	0.096	0.295	37045
Senior secondary education	0.331	0.471	37045
Diploma	0.104	0.305	37045
Graduate degree, non-STEM	0.270	0.444	37045
Graduate degree, STEM	0.152	0.359	37045
Postgraduate degree, non-STEM	0.027	0.162	37045
Postgraduate degree, STEM	0.018	0.134	37045
Other (education not specified)	0.001	0.038	37045
Experience:			
$0 - 1$ years	0.504	0.500	37020
$1 - 2$ years	0.336	0.472	37020
$2 - 5$ years	0.037	0.188	37020
> 5 years	0.123	0.329	37020
Other job characteristics:			
Non-metro location	0.405	0.491	37045
Number of applications	44.779	127.208	37045
Wage not specified	0.439	0.496	37045
Annual wage (Rs.), if wage specified in job ad	240665.174	179959.717	20784

Notes: Wages and experience are the mid-point of the range specified in the job ad.

Table A.2: Descriptive statistics: Job applicants

	Male	Female	Total
Education:			
Secondary education	0.012	0.005	0.010
Senior secondary education	0.084	0.048	0.071
Diploma	0.073	0.027	0.057
Graduate degree, STEM	0.539	0.512	0.529
Graduate degree, non-STEM	0.144	0.174	0.155
Postgraduate degree, STEM	0.060	0.103	0.075
Postgraduate degree, non-STEM	0.085	0.128	0.100
$>$ Postgraduate	0.001	0.002	0.001
Other	0.002	0.002	0.002
Experience:			
<1 years	0.726	0.787	0.748
$1-2$ years	0.082	0.078	0.080
$2-5$ years	0.107	0.090	0.101
>5 years	0.085	0.046	0.071
Other candidate characteristics			
Candidate age	24.648	23.941	24.395
Non-metro location	0.579	0.541	0.566
Number of applications	3.824	4.029	3.897
N (Applicants)	145723	80741	226464

Notes: Each cell gives the average value of the variable in the respective sub-sample of job applications. Experience is given in years and is divided into four categories to correspond to the job advertisements sample.
Source: The applicant sample includes those who applied to at least one job in our job advertisement sample and disclosed their gender.

Table A.3: Descriptive statistics: Job ads (Shortlisting data)

	Mean	**SD**	**N**
Education:			
Secondary education or less	0.114	0.318	888
Senior secondary education	0.579	0.494	888
Diploma	0.066	0.249	888
Graduate degree, non-STEM	0.188	0.391	888
Graduate degree, STEM	0.048	0.215	888
Postgraduate degree, non-STEM	0.003	0.058	888
Postgraduate degree, STEM	0.001	0.034	888
Other (education not specified)	0.000	0.000	888
Experience:			
$0-1$ years	0.918	0.275	888
$1-2$ years	0.048	0.215	888
$2-5$ years	0.005	0.067	888
>5 years	0.029	0.169	888
Other job characteristics:			
Non-metro location	0.170	0.376	888
Number of applications	63.525	160.546	888
Wage not specified	0.002	0.047	888
Annual wage (Rs.), if wage specified in job ad	255546.660	180067.321	886

Notes: Wages and experience are the mid-point of the range specified in the job ad. The set of job ads for which shortlisting data by employers is available

B Additional Figures and Tables

Table B.1: Effect of Distance on Application by candidates: Heterogeneity by posted wages

	(1)	(2)	(3)	(4)	(5)
1-50	–0.038***	–0.383***	–0.038***	–0.044***	–0.058***
	(0.002)	(0.042)	(0.002)	(0.003)	(0.005)
50-100	–0.090***	–0.372***	–0.090***	–0.096***	–0.109***
	(0.001)	(0.016)	(0.001)	(0.001)	(0.003)
100-500	–0.107***	–0.379***	–0.107***	–0.112***	–0.133***
	(0.001)	(0.010)	(0.001)	(0.001)	(0.002)
> 500	–0.113***	–0.373***	–0.113***	–0.117***	–0.134***
	(0.001)	(0.010)	(0.001)	(0.001)	(0.002)
ln (wage)	0.006***				
	(0.000)				
$\mathbb{1}$(wage $>$ med)			0.006***		
			(0.000)		
1-50 $\times$ ln (wage)		0.028***			
		(0.003)			
50-100 $\times$ ln (wage)		0.023***			
		(0.001)			
100-500 $\times$ ln (wage)		0.022***			
		(0.001)			
> 500 $\times$ ln (wage)		0.022***			
		(0.001)			
1-50 $\times$ $\mathbb{1}$(wage $>$ med)				0.022***	0.023***
				(0.004)	(0.004)
50-100 $\times$ $\mathbb{1}$(wage $>$ med)				0.020***	0.019***
				(0.001)	(0.001)
100-500 $\times$ $\mathbb{1}$(wage $>$ med)				0.021***	0.021***
				(0.001)	(0.001)
> 500 $\times$ $\mathbb{1}$(wage $>$ med)				0.020***	0.020***
				(0.001)	(0.001)
Constant	0.048***	0.114***	0.121***	0.113***	0.117***
	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)
N	62515398	62575521	62515398	62575521	51419446
Controls					
Job controls	✓		✓		
Job FE		✓		✓	✓
Candidate FE	✓	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓	✓
Cand. controls x distance					✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Interaction controls include distance interacted for graduates above and experience of more than one year for candidates. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with wage information available location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table B.2: Effect of Distance on Application by candidates

	(1)	(2)	(3)
1-50	–0.013***	–0.028***	–0.027***
	(0.002)	(0.003)	(0.004)
50-100	–0.048***	–0.072***	–0.070***
	(0.001)	(0.001)	(0.002)
100-500	–0.057***	–0.086***	–0.084***
	(0.000)	(0.001)	(0.001)
> 500	–0.061***	–0.087***	–0.085***
	(0.000)	(0.001)	(0.001)
1-50 x Female	–0.008***	–0.008***	–0.006*
	(0.003)	(0.003)	(0.003)
50-100 x Female	–0.008***	–0.010***	–0.011***
	(0.001)	(0.001)	(0.001)
100-500 x Female	–0.004***	–0.007***	–0.008***
	(0.001)	(0.001)	(0.001)
> 500 x Female	–0.004***	–0.006***	–0.007***
	(0.001)	(0.001)	(0.001)
1-50 x Graduate		0.017***	0.015***
		(0.004)	(0.004)
50-100 x Graduate		0.030***	0.029***
		(0.001)	(0.002)
100-500 x Graduate		0.034***	0.033***
		(0.001)	(0.001)
> 500 x Graduate		0.032***	0.031***
		(0.001)	(0.001)
1-50 x Experienced			0.000
			(0.004)
50-100 x Experienced			–0.011***
			(0.001)
100-500 x Experienced			–0.011***
			(0.001)
> 500 x Experienced			–0.011***
			(0.001)
Constant	0.065***	0.065***	0.067***
	(0.000)	(0.000)	(0.000)
N	166555050	166415759	135272680
Job FE	✓	✓	✓
Candidate FE	✓	✓	✓
Month-year FE	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table B.3: Effect of Distance on Application (Graduates and above candidates): Heterogeneity by metro vs non-metro job location

Sample:	All jobs	Jobs with posted wages		
	(1)	(2)	(3)	(4)
1-50	–0.013***	–0.021***	–0.033***	–0.020***
	(0.002)	(0.003)	(0.004)	(0.004)
50-100	–0.030***	–0.050***	–0.059***	–0.052***
	(0.000)	(0.001)	(0.001)	(0.001)
100-500	–0.038***	–0.063***	–0.071***	–0.065***
	(0.000)	(0.000)	(0.001)	(0.001)
> 500	–0.043***	–0.072***	–0.080***	–0.074***
	(0.000)	(0.001)	(0.001)	(0.001)
1-50 x 𝟙(Non-metro)	–0.037***	–0.102***	–0.101***	–0.145***
	(0.003)	(0.005)	(0.005)	(0.006)
50-100 x 𝟙(Non-metro)	–0.063***	–0.136***	–0.136***	–0.168***
	(0.001)	(0.002)	(0.002)	(0.003)
100-500 x 𝟙(Non-metro)	–0.063***	–0.141***	–0.141***	–0.173***
	(0.001)	(0.002)	(0.002)	(0.002)
> 500 x 𝟙(Non-metro)	–0.064***	–0.139***	–0.139***	–0.170***
	(0.001)	(0.002)	(0.002)	(0.002)
1-50 x 𝟙(wage> med)			0.023***	–0.002
			(0.004)	(0.005)
50-100 x 𝟙(wage> med)			0.017***	0.003**
			(0.001)	(0.002)
100-500 x 𝟙(wage> med)			0.017***	0.004***
			(0.001)	(0.001)
> 500 x 𝟙(wage> med)			0.016***	0.004***
			(0.001)	(0.001)
1-50 x 𝟙(wage> med) x 𝟙(Non-metro)				0.093***
				(0.009)
50-100 x 𝟙(wage> med) x 𝟙(Non-metro)				0.066***
				(0.004)
100-500 x 𝟙(wage> med) x 𝟙(Non-metro)				0.064***
				(0.003)
>500 x 𝟙(wage> med) x 𝟙(Non-metro)				0.063***
				(0.003)
Constant	0.071***	0.151***	0.140***	0.141***
	(0.000)	(0.001)	(0.001)	(0.001)
N	151665025	54491125	54491125	54491125
Mean Y	0.009	0.017	0.017	0.017
Mean Y (Same city)	0.0608	0.107	0.107	0.107
Controls				
Job FE	✓	✓	✓	✓
Candidate FE	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table B.4: Effect of Distance on Application by candidates: Robustness

	(1)	(2)	(3)
1-50	–0.013***	–0.023***	–0.022***
	(0.001)	(0.002)	(0.004)
50-100	–0.036***	–0.058***	–0.058***
	(0.000)	(0.001)	(0.001)
100-500	–0.041***	–0.070***	–0.069***
	(0.000)	(0.001)	(0.001)
$>$ 500	–0.046***	–0.074***	–0.078***
	(0.000)	(0.001)	(0.001)
1-50 x $\mathbb{1}$(Non-metro)	–0.059***	–0.061***	–0.156***
	(0.003)	(0.003)	(0.006)
50-100 x $\mathbb{1}$(Non-metro)	–0.077***	–0.078***	–0.174***
	(0.001)	(0.001)	(0.002)
100-500 x $\mathbb{1}$(Non-metro)	–0.082***	–0.081***	–0.180***
	(0.001)	(0.001)	(0.002)
$>$ 500 x $\mathbb{1}$(Non-metro)	–0.084***	–0.084***	–0.178***
	(0.001)	(0.001)	(0.002)
1-50 x $\mathbb{1}$(wage$>$ med)			–0.002
			(0.004)
50-100 x $\mathbb{1}$(wage$>$med)			0.006***
			(0.001)
100-500 x $\mathbb{1}$(wage$>$ med)			0.008***
			(0.001)
$>$ 500 x $\mathbb{1}$(wage$>$ med)			0.008***
			(0.001)
1-50 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)			0.083***
			(0.008)
50-100 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)			0.070***
			(0.003)
100-500 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)			0.066***
			(0.003)
$>$500 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)			0.065***
			(0.003)
Constant	0.082***	0.084***	0.144***
	(0.000)	(0.000)	(0.001)
N	166554699	166487070	62565076
Candidate FE	✓	✓	✓
Job FE	✓	✓	✓
Month-year FE	✓	✓	✓
Candidate FE $\times$ Non-metro	✓	✓	✓
Job edu, exp $\times$ distance		✓	
Candidate FE $\times$ $\mathbb{1}$(wage$>$ med)			✓
Candidate FE $\times$ $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)			✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with wage information available location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table B.5: Effect of Distance on Application by candidates: Robustness

	(1)	(2)	(3)	(4)
1-50	–0.394***			
	(0.042)			
50-100	–0.320***			
	(0.015)			
100-500	–0.286***			
	(0.009)			
$>$ 500	–0.283***			
	(0.009)			
1-50 x $\mathbb{1}$(Non-metro)	–0.115***	–0.100***	–0.087***	–0.102***
	(0.004)	(0.003)	(0.004)	(0.008)
50-100 x $\mathbb{1}$(Non-metro)	–0.145***	–0.122***	–0.120***	–0.135***
	(0.002)	(0.002)	(0.002)	(0.004)
100-500 x $\mathbb{1}$(Non-metro)	–0.153***	–0.126***	–0.128***	–0.143***
	(0.002)	(0.001)	(0.001)	(0.003)
$>$ 500 x $\mathbb{1}$(Non-metro)	–0.152***	–0.128***	–0.129***	–0.145***
	(0.002)	(0.001)	(0.001)	(0.003)
Constant	0.148***	0.299	0.054	–0.156
	(0.001)	(34657.891)	(16531.265)	(14635.43)
N	62514607	62514607	62513819	62208200
Candidate FE	✓	✓	✓	✓
Job FE	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓
Candidate FE × Non-metro	✓	✓	✓	✓
Job edu, exp × distance	✓	✓	✓	✓
Job role × distance		✓		
Job title × distance			✓	
Job title, CompanyID × distance				✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Interaction controls include distance interacted for graduates above and experience of more than one year for candidates. Additional controls include Job education, experience requirement × distance. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with wage information available location in a single city and job role excluding 'others', for which any candidates applied between July 2020- December 2020.

Table B.6: Effect of Distance on Application by candidates (2019)

	(1)	(2)	(3)	(4)
1-50	–0.005***	–0.006***	–0.007***	–0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
50-100	–0.028***	–0.030***	–0.032***	–0.031***
	(0.000)	(0.000)	(0.000)	(0.000)
100-500	–0.032***	–0.034***	–0.035***	–0.035***
	(0.000)	(0.000)	(0.000)	(0.000)
> 500	–0.034***	–0.036***	–0.038***	–0.037***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.036***	0.039***	0.043***	0.038***
	(0.000)	(0.002)	(0.001)	(0.000)
N	162506102	161293860	162497861	162505938
Mean Y	0.006	0.006	0.006	0.006
Mean Y (Same city)	0.0356	0.0356	0.0356	0.0356
Controls				
Job controls			✓	
Candidate controls		✓		
Job FE		✓		✓
Candidate FE			✓	✓
Month-year FE	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between October 2019- December 2019.

Table B.7: Effect of Distance on Application by candidates (2019): Metro vs non-metro

Sample:	All jobs	Jobs with posted wages		
	(1)	(2)	(3)	(4)
1-50	–0.010***	–0.015***	–0.016***	–0.016***
	(0.001)	(0.002)	(0.003)	(0.003)
50-100	–0.022***	–0.033***	–0.035***	–0.032***
	(0.000)	(0.001)	(0.001)	(0.001)
100-500	–0.024***	–0.039***	–0.041***	–0.039***
	(0.000)	(0.000)	(0.001)	(0.001)
> 500	–0.029***	–0.044***	–0.045***	–0.043***
	(0.000)	(0.000)	(0.001)	(0.001)
1-50 x $\mathbb{1}$(Non-metro)	–0.019***	–0.037***	–0.037***	–0.042***
	(0.002)	(0.004)	(0.004)	(0.006)
50-100 x $\mathbb{1}$(Non-metro)	–0.040***	–0.067***	–0.067***	–0.077***
	(0.001)	(0.001)	(0.001)	(0.002)
100-500 x $\mathbb{1}$(Non-metro)	–0.041***	–0.067***	–0.067***	–0.076***
	(0.001)	(0.001)	(0.001)	(0.002)
> 500 x $\mathbb{1}$(Non-metro)	–0.040***	–0.066***	–0.067***	–0.076***
	(0.001)	(0.001)	(0.001)	(0.002)
1-50 x $\mathbb{1}$(wage$>$ med)			0.002	0.002
			(0.003)	(0.004)
50-100 x $\mathbb{1}$(wage$>$ med)			0.003***	–0.001
			(0.001)	(0.001)
100-500 x $\mathbb{1}$(wage$>$ med)			0.002***	–0.001*
			(0.001)	(0.001)
> 500 x $\mathbb{1}$(wage$>$ med)			0.001	–0.002***
			(0.001)	(0.001)
1-50 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)				0.009
				(0.007)
50-100 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)				0.016***
				(0.003)
100-500 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)				0.015***
				(0.002)
$>$500 x $\mathbb{1}$(wage$>$ med) x $\mathbb{1}$(Non-metro)				0.016***
				(0.002)
Constant	0.047***	0.080***	0.075***	0.075***
	(0.000)	(0.001)	(0.000)	(0.001)
N	162505938	78596823	78596823	78596823
Mean Y	0.006	0.009	0.009	0.009
Mean Y (Same city)	0.0356	0.0578	0.0578	0.0578
Controls				
Job FE	✓	✓	✓	✓
Candidate FE	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate applied for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Regressions weighted by inverse of total applications by each candidate. Robust standard errors clustered at the candidate level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any candidates applied between October 2019- December 2019.

Table B.8: Effect of Distance on Shortlisting by Employer: Heterogeneity

	(1)	(2)	(3)	(4)	(5)
1-50	0.007	–0.019	–0.011	0.013	–0.060
	(0.032)	(0.028)	(0.032)	(0.089)	(0.105)
50-100	0.057**	0.022	0.041	–0.050	0.127
	(0.026)	(0.023)	(0.027)	(0.081)	(0.184)
100-500	0.018	0.014	0.013	–0.015	–0.061
	(0.012)	(0.011)	(0.013)	(0.027)	(0.079)
> 500	–0.031**	–0.031**	–0.027**	–0.034	–0.064
	(0.013)	(0.012)	(0.013)	(0.024)	(0.057)
Female	–0.019**	–0.017**	–0.014		
	(0.008)	(0.008)	(0.010)		
1-50 x Female			–0.028	–0.027	–0.027
			(0.046)	(0.103)	(0.101)
50-100 x Female			–0.061	0.078	0.090
			(0.042)	(0.113)	(0.107)
100-500 x Female			0.007	0.048	0.045
			(0.020)	(0.043)	(0.043)
> 500 x Female			–0.010	–0.060	–0.062
			(0.017)	(0.039)	(0.039)
Constant	0.342**	0.211	0.209	0.324***	0.324***
	(0.173)	(0.135)	(0.135)	(0.008)	(0.009)
N	55739	55738	55738	28158	28133
Mean Y	0.295	0.295	0.295	0.295	0.295
Mean Y (Same city, base)	0.357	0.357	0.372	0.372	0.413
Wald test: Gender			0.031	0.004	0.052
Controls					
Candidate controls	✓	✓	✓		
Cand. edu × Distance					✓
Candidate FE				✓	✓
Job controls	✓				
Job FE		✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓	✓

Notes: The dependent variable takes a value of one if a candidate was shortlisted by an employer for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Candidate controls include gender, education, age and age square and city location of the candidate. Regressions weighted by inverse of total applicants to a posted job. Each column reports the effective number of observations after incorporating the included fixed effects. Robust standard errors clustered at the job level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city and job role excluding 'others', for which any shortlisting was done by the employers during July 2020 - December 2020.

Table B.9: Effect of Distance on Shortlisting by Employer: Heterogeneity (Robustness)

	(1)	(2)	(3)	(4)	(5)	(6)
1-50	–0.014	0.066	0.046	0.034	0.071	0.063
	(0.021)	(0.056)	(0.073)	(0.037)	(0.097)	(0.098)
50-100	0.059**	–0.028	0.100	–0.012	–0.044	0.009
	(0.025)	(0.059)	(0.132)	(0.032)	(0.066)	(0.066)
100-500	0.029**	0.033	0.050	–0.015	–0.035	–0.027
	(0.013)	(0.024)	(0.049)	(0.016)	(0.026)	(0.029)
> 500	–0.014	0.005	0.013	–0.063***	–0.093***	–0.078***
	(0.012)	(0.021)	(0.040)	(0.014)	(0.023)	(0.026)
Female	–0.019**			–0.023***		
	(0.009)			(0.007)		
1-50 x Female	–0.054	–0.197*	–0.204*			
	(0.037)	(0.117)	(0.116)			
50-100 x Female	–0.144***	0.021	0.033			
	(0.041)	(0.104)	(0.102)			
100-500 x Female	–0.005	–0.005	–0.002			
	(0.019)	(0.041)	(0.041)			
> 500 x Female	0.015	–0.086**	–0.086**			
	(0.016)	(0.037)	(0.037)			
1-50 x $\mathbb{1}$(Non-metro)				–0.075	–0.051	–0.057
				(0.053)	(0.133)	(0.130)
50-100 x $\mathbb{1}$(Non-metro)				0.064	0.048	0.069
				(0.043)	(0.129)	(0.114)
100-500 x $\mathbb{1}$(Non-metro)				0.078***	0.153***	0.146***
				(0.024)	(0.042)	(0.042)
> 500 x $\mathbb{1}$(Non-metro)				0.156***	0.228***	0.232***
				(0.026)	(0.040)	(0.040)
Constant	0.382***	0.360***	0.356***	0.390***	0.380***	0.375***
	(0.103)	(0.007)	(0.008)	(0.103)	(0.010)	(0.010)
N	71489	40922	40856	71489	40922	40922
Wald test: Gender	0.984	0.010	0.132			
Controls						
Candidate controls	✓			✓		
Job FE	✓	✓	✓	✓	✓	✓
Candidate FE		✓	✓		✓	✓
Month-year FE	✓	✓	✓	✓	✓	✓
Cand. Edu, Exp x Distance			✓			
Job Edu, Exp × Distance						✓

Notes: The dependent variable takes a value of one if a candidate was shortlisted by an employer for a job and zero otherwise. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. All specifications control for month and year of job posting. Candidate controls include gender, education, age and age square and city location of the candidate. Additional controls include Candidate Edu, Exp × Distance. Regressions weighted by inverse of total applicants to a posted job. Each column reports the effective number of observations after incorporating the included fixed effects. Robust standard errors clustered at the job level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Data includes all full time job ads posted on the platform with location in a single city, for which any shortlisting was done by the employers during July 2020 - December 2020.