Job Search: Bridging the Miles to Non-Metro Locations*

Kanika Mahajan† Vandana Saini‡ Shekhar Tomar§

Ashoka University *IIM-B* *ISB*

February 28, 2025

Abstract

Using data from a large online job portal, we examine how distance and city size impact job seekers' application behavior. We find that candidates have a relative preference for larger (metro) cities, even when they are located far from candidates' home cities, while they have a strong distaste to non-metro cities, with application rates dropping to near zero for jobs located more than 100 miles away. Interestingly, employers in non-metro cities do not exhibit similar distaste for distant candidates. Leveraging a natural experiment, we find that enhanced transportation connectivity does not alleviate this aversion to non-metro locations. However, non-metro cities with dense labor markets for specific job roles help mitigate this disparity.

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

Keywords: Job search, Distaste for non-metro, Air Route, Job Concentration

*We thank Shilpa Aggarwal, Shashwat Alok, Bhagwan Chaudhary, Nikhil Datta, Robert Garlick, Krishna B. Kumar, Dilip Mookherjee, Roland Rathelot, Marcus Roesch, and conference participants at IEG, ISB, Jobs and Development Conference (Cape Town, 2022), Indian Statistical Institute, Urban Economics Association (Toronto, 2023) for helpful comments and suggestions. 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 Institute of Management, Bengaluru, India. Email: vandanas23@iimb.ac.in

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

1 Introduction

Domestic migration facilitates economic growth and urbanization, and exhibits wide variation across cities. For instance, cities that are far away from a job seekers' location (Porcher, 2020; Marinescu & Rathelot, 2018) and relatively smaller in size (De la Roca *et al.*, 2023a) often attract fewer skilled workers. The latter distinction is even more pronounced in developing economies, where workers tend to cluster in a handful of large cities. While the limited number of jobs in smaller cities is often cited as a major factor, it remains unclear whether this is compounded by a lower preference of workers to move to these cities. Understanding the interplay between distance, city size, and job seekers' preferences is crucial for shaping urban development, resource allocation, and economic policy as these countries continue to grow. In this paper, we study how distance and city size interact to shape spatial job search behavior. We first document a novel phenomenon - job seekers prefer larger vs. smaller cities farther from their city of residence, and then assess the role of place-based policies that decrease migration costs in mitigating this differential preference. Our findings reveal that only smaller cities with dense labor markets, i.e., those with high density of jobs within a given sector, can compete with larger cities in attracting job seekers.

We use novel data set from a major job portal for entry level job seekers that includes job advertisements, applications, and shortlisting information between 2018 and 2020 in India. The portal primarily caters to young, and first-time job seekers – a relatively mobile group (Global Migration Group, Amior (2024)), making it ideal for studying spatial job preferences. Additionally, India provides an ideal context for this investigation for several reasons. First, India is currently pre-dominantly rural (65 percent) but is expected to urbanize rapidly. This transition offers an excellent background to study what makes certain cities more attractive during periods of rapid economic growth.[1] Second, India's urban growth is highly concentrated, with the largest (or metro) cities exhibiting far higher population

[1] Chauvin *et al.* (2017) argue that countries like India have not achieved spatial equilibrium yet and understanding urbanization in developing economies can provide important insights into city growth.

densities than their counterparts in developed economies like the US or emerging economies like China (Desmet *et al.*, 2015; Chauvin *et al.*, 2017). Consequently, understanding why workers gravitate overwhelmingly towards metro cities is important for urban planning and firm location decisions.[2]

Using data from 37,000 job ads for full-time positions and 226,000 candidates who applied to these vacancies, we first document whether there is a differential preference for jobs across distant cities by their size. We observe job characteristics, such as job role, city location, education and experience requirements, company ID, and posted wages. 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. For a subset of jobs, we also observe information on shortlisting by firms.

To analyze spatial job search behavior, we construct a detailed choice set for each candidate based on job role and application date. It gives us 166 million observations at candidate×job ad level, with an average application rate of 1 percent.[3] The application rate 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 with an increase in distance between the candidate and job locations, similar to the effect of distance in migration gravity models (Grogger & Hanson, 2011; Bryan & Morten, 2015). Quantitatively, the application rate for a job located 100 miles away, relative to the candidate's home city, falls by 90 percent and approaches zero for distances above 500 miles.[4]

Crucially, the application rate significantly differs for distant jobs located in the largest metropolitan (metro) cities vs. other (non-metro) cities.[5] We find that the distaste for

[2]Rapid urbanization can also cause over-crowding, congestion, and proliferation of slums, leading to subsequent deterioration of basic services, health, and poverty outcomes (see survey by (Marx *et al.*, 2013)).

[3]He *et al.* (2021) experimentally find a similar application rate, 0.5%, for a job board in China.

[4]We find that candidate attributes, like gender, education, and experience play a minor role in explaining the magnitude. The baseline magnitude for distaste in our sample is intermediate relative to Marinescu & Rathelot (2018) and Banfi *et al.* (2019), the studies that estimate the distaste using applications data.

[5]For classification purposes, cities are grouped based on their urban agglomeration populations as reported in the most recent Indian Census of 2011. According to this demarcation, Delhi, Mumbai, Bengaluru, Chennai, and Kolkata emerge as the principal metro cities (top 5 based on population with each having at least 8 million residents). For a comprehensive list of Indian cities by urban population, see Census 2011.

distance (beyond 100 miles) is largely driven by jobs in non-metro cities with distaste for these locations three times as high as that for metros. At 500 miles, candidates are one-third as likely to apply to metro jobs compared to those in their home city. In contrast, the likelihood of applying to non-metro jobs beyond 500 miles falls to nearly zero. These findings remain stable with stricter controls (e.g. identical roles within the same firm), posted wages, candidate ability, or using city population instead of metro/non-metro classification. Therefore, our results are not driven by job-specific attributes that differ across cities and influence candidate's job search behavior. Our regression estimates suggest that workers would require a wage 3.75 times the median wage to equate the application rate for a non-metro city 500 miles away with that of their home city.

Subsequently, we explore some factors that can potentially explain the disparity in candidate preference for metro vs. non-metro locations. We first assess whether place-based policies focusing on infrastructure investments to enhance transportation connectivity between cities could reduce distaste for non-metro locations. Typically, metro cities offer better connectivity to the rest of the country, and may explain why job seekers prefer them (through lower migration costs and ease of coming back to home city). If connectivity is a significant factor in explaining the divergence between metro vs. non-metro cities, reducing connectivity costs (both pecuniary and non-pecuniary) could potentially make non-metro cities more appealing.

To investigate this, we utilize the natural experiment provided by the *UDAN* scheme, which introduced new subsidized air routes across Indian cities.[6] Implementing a staggered difference-in-differences estimation strategy, we compare application rates from city pairs connected under the *UDAN* scheme with those that were not, controlling for job attributes. We find no significant change in the average number of applications to jobs located in non-metro cities after *UDAN*. However, applications to jobs in metro cities increase by 25 percent. These results suggest that while improved connectivity further enhances the attractiveness of

[6]UDAN stands for "Ude Desh ka Aam Naagrik" which translates to "Let the common citizen of the country fly." Section 4.1 and Appendix Section B.3 provide details on the scheme.

metro cities, it does little to alleviate the distaste for non-metro locations, suggesting that connectivity is not the primary factor driving candidate preference for larger cities.

Next, we examine the role of job role concentration in cities on the observed divergence in distaste across metro vs. non-metro cities. Having a high density of jobs in a particular sector offers workers greater opportunities for employer switching and enhancing career prospects within the same city. In our context, non-metro cities functioning as industrial clusters might attract candidates seeking dynamic advantages for career growth in specific sectors.[7] For example, a candidate starting a career in the pharmaceutical industry might find Hyderabad an ideal location, while someone aiming for a career in finance might not.[8]

To investigate this, we analyze application behavior as a function of dynamic concentration of job role in a city. We exploit the timing of job search spells by a candidate, since most candidates appear for a short time frame for searching jobs on the platform, to calculate the proportion of jobs in a given job role posted across cities and visible to a candidate on the portal. This dynamic measure allows identification of distaste arising from variation in job role concentration based on timing of job search spells and filters out any city-job role specific non-time-varying attributes. We find that a higher concentration of a given job role in a city significantly reduces the distaste for that job role even when the city is located far away. Notably, for the non-metro locations the distaste completely disappears beyond 100 miles if the city contributes 50 percent postings for a given job role. These results suggest that job role density can mitigate some of the distaste for non-metro locations.

Furthermore, we examine the interaction between transportation connectivity ($UDAN$) and job role density in shaping candidate application behavior. Interestingly, we find that non-metro cities with an above-median concentration of jobs in a specific role experience a significant increase in applications after being connected by an $UDAN$ route. This suggests that while improved connectivity alone is insufficient to make non-metro cities attractive, it

[7]This is similar to the job surplus hypothesis in Amior (2024), who argues that job surplus in destination city matters for long distance migration.

[8]Hyderabad hosts a vibrant pharmaceutical sector and accounts for one-third of the global vaccine production and 30 percent of India's pharmaceutical production (see Invest Telangana).

acts as an effective amplifier for cities already established as hubs for specific job roles.

Finally, we examine whether non-metro employers exhibit similar distaste for distant candidates. If yes, then the observed candidate distaste is an equilibrium outcome of job search process where candidates do not apply to distant non-metro jobs as employers are unlikely to hire them. Accounting for job and candidate level unobservables, we find that employers in non-metro locations are 11 percentage points more likely to consider distant candidates for the job, effectively offsetting the baseline distaste for distant applicants. On average, jobs in non-metro locations receive one-third fewer applications relative to the mean application rate, which may explain why employers in these locations display no distaste for distant candidates.[9] Taken together, these results show that the distaste for distance non-metro cities by job seekers is not explained by them facing a lower probability of selection by employers.

Our work contributes to the literature on labor mobility by exploring how agglomeration forces and city characteristics shape workers' preferences, particularly their distaste to move to smaller cities. Prior studies have highlighted that workers choose big cities to have higher earnings (Roca & Puga, 2017; Eeckhout *et al.*, 2014; Baum-Snow & Pavan, 2012) and access better amenities like education (De la Roca *et al.*, 2023b; Arntz *et al.*, 2023; Glaeser *et al.*, 2001), among others. De la Roca *et al.* (2023a) shows that younger workers having higher self-belief are more likely to migrate to bigger cities. This paper adds to the existing literature by first documenting a higher preference of workers for distant larger cities, even after controlling for posted wages and job and candidate level unobservables. Our estimates suggest that jobs in non-metro locations receive almost no application from candidates located 100 miles away, quantifying the extent of candidate distaste for non-metro locations. We then provide causal evidence on two key aspects. First, while improved transport connectivity enhances accessibility, our findings suggest that it does not reduce the distaste for non-metro locations. Second, dense labor markets for specific job roles in non-metro cities can partially mitigate

[9] In a recent survey conducted by Mint and Shine, over 200 human resource executives expressed concerns about the challenges of recruiting skilled workers in smaller cities.

this distaste.

Our result on dense labor markets overlaps with the literature on industrial cluster policies. The evidence on the effectiveness of such policies remains sparse and inconclusive (see survey by Neumark & Simpson (2015)).[10] However, none of these studies examine whether such policies affect job search behavior or migration. The only study to examine the role of place-based tax-exemption policy on migration is by Abeberese *et al.* (2024) who find an increase in employment driven by migration into the treated districts. Utilizing postings data to construct time-varying occupation concentration across cities, exploiting exogenous variation in candidate search spells, we demonstrate a preference among job seekers for cities with dense labor markets. These results suggest that small cities could be attractive for job seekers if they have dense labor markets, which can perhaps provide an enabling environment to start a virtuous cycle for local economic growth. Moreover, given that we find that employers in non-metro cities do not exhibit any distaste, policies aimed at encouraging firms (via subsidies) in non-metro areas to increase employment may not effectively attract more job seekers unless there is a big push to create substantive number of jobs in the city.

More broadly, our *UDAN*-based results connect to the literature on evaluating the economic impact of airport expansions. Blonigen & Cristea (2015) and McGraw (2020) find that the expansion of airports increased economic and population growth in connected cities in the US, and Gibbons & Wu (2020) observe similar impact on manufacturing activity in China. Conversely, Sheard (2014) and Breidenbach (2020) find no significant economic impact of airport extensions in the US and Germany, respectively.[11] Relative to this literature, our

[10]While Falck *et al.* (2010) find positive impact on innovation in targeted sectors in Germany, Huber (2012); Martin *et al.* (2011); McDonald *et al.* (2007) report no significant benefits for local economy. Similarly, Freedman *et al.* (2023) find no impact of Opportunity Zone policy on employment, while Criscuolo *et al.* (2019) find positive impact of investment subsidy on employment within small firms. This aligns with the evidence in Atkins *et al.* (2023) who find limited impact of Opportunity Zone program, that grants tax breaks for investment in designated areas, on job postings. For a theoretical exploration of hiring incentives in distressed regions, refer to Kline & Moretti (2013). A substantial body of empirical research examines the efficacy of place-based policies in distressed areas (see survey by (Bartik, 2020)). For research on developed economies, see (Hyman *et al.*, 2022; Mayer *et al.*, 2017; Busso *et al.*, 2013; Ham *et al.*, 2011; Neumark & Kolko, 2010; Bondonio & Greenbaum, 2007), and for developing economies, see (Hasan *et al.*, 2021; Lu *et al.*, 2019; Chaurey, 2017; Wang, 2013).

[11]This also relates to the broader research evaluating the economic impacts of transportation infrastructure

paper uniquely documents the influence of air connectivity on job search process and how it differs based on job location. We find heterogeneity in gains as large metro cities benefit from new air routes by further attracting more job seekers, suggesting a siphoning effect. We also contribute to the nascent work on understanding urbanization in developing countries. While agglomeration forces continue to work, findings from advanced economies do not translate one-to-one to these countries (Chauvin *et al.*, 2017). This distinction underscores the need for evidence-based research to inform urbanization policies (Bryan *et al.*, 2020; Moretti, 2014). However, lack of quality data has so far impeded such research work. By leveraging job search data, our study provides causal insights into the factors that draw job seekers to cities, enhancing our understanding of city development.

The remainder of the paper is organized as follows. Section 2 discusses the data and Section 3 quantifies job seeker distaste for distant jobs located in non-metro. Section 4 evaluates the channels that contribute towards the distaste for jobs indistant non-metro cities. Section 5 concludes.

2 Data

We use data from one of the major online job platforms catering 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.[12] The final hiring decision, however, is not observed by the portal. For analysing the distaste for distant jobs and specifically how this varies by city size, we draw from data related to job ads posted from July–December 2020, and the applications

projects like highways and railways on employment or urban growth (Duranton & Turner, 2012; Lin, 2017), urban firms (Gibbons *et al.*, 2019; Baum-Snow, 2007), skill premiums (Michaels, 2008), long-term economic growth (Banerjee *et al.*, 2020) and trade (Donaldson, 2018). Also see Redding & Turner (2015) for a survey on the impact of transportation networks.

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

made to these jobs. We choose this time frame for our job ad level analyses since for a subset of these jobs ads, data on shortlisting by the employers, is also available. Thus, we observe information on both sides of the matching market during this time period.[13] We also have data on posted ads and applications from June 2018–December 2019 which we use to examine the role played by policies that reduce connectivity costs in affecting applications (Section 4). We also use it to test robustness of our main findings on distaste in alternate periods.

For the posted jobs, we have extensive information on variables such as date of posting, job role, job title, education requirements, minimum and maximum required experience, minimum and maximum offered wage, job location, firm ID 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 apart from "Others". After removing ads that have no applications within the first ten days of posting, we have a final sample of 37,045 job ads across 419 cities. We elaborate the importance of this last step in the next section.

The descriptive statistics for these jobs ads are provided in Appendix Table A.1. As mentioned earlier, these jobs largely cater to young job seekers, with 50% and 33% jobs posting a requirement of 0-1 and 1-2 years of experience, respectively. Notably, around 56% of the jobs ads in our data post wages, which is larger than many other existing studies that use posted wages. However, this is similar to data from other job platforms in India (Chaurey *et al.*, 2024; Chaturvedi *et al.*, 2024).[14] 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 data, are significant economic and population hubs. For

[13]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.

[14]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).

instance, Delhi and Mumbai are the largest cities in north and west India, respectively. Appendix Figure 1 shows the distribution of cities across the country for the posted jobs.

The second dataset provides 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 2).[15] Approximately 44% of them reside in metro locations. We provide descriptive statistics for candidates in Appendix Table A.2. The third data on applications, allows us to map candidate applications to the posted job ads. On average, candidates apply to 4 jobs and each job ad receives 45 applications.

Finally, the fourth dataset provides information on shortlisting. We observe the candidates shortlisted 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 data, we keep the 888 job ads that do not mention "Others" in their job 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.

Appendix Section B.1 provides more detailed information on these data sets, and the specifics of data processing.

3 Distaste for Distant Jobs in Non-metro

In this section, we examine candidate's distaste for distant job opportunities and whether it varies across distant metro vs. non-metro cities. We first describe the construction of the choice set, followed by the estimation strategy and the results.

[15] 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 less serious candidates from our analyses.

3.1 Constructing Candidate Choice Sets

To investigate how the geographical distance between job and candidate location and its interaction with city size (metro vs. non-metro) influences application rates, we require job choice sets for candidates. While we observe the set of job ads that candidates apply to, we do not observe the jobs they choose to forego. Consequently, we need to construct choice sets of potential jobs for the applicants. 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 portal sorts job ads by their posting date as default. Our discussions with the platform reveal that usually candidates filter job ads based on their preferred roles and then the displayed ads are sorted by the posting date. Thus, more recent job ads are listed at the top and have a higher likelihood of being seen by a candidate within her set of preferred job roles. This is corroborated in the data as well. Figure 3, 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 the highest during the first two days (7 per day) and then declines 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, justifying 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. These patterns in search behavior are consistent with findings in (Davis & de la Parra, 2017) for the US, suggesting that candidate job search spells are usually of short durations.

This approach yields a choice set of 166 million candidates$\times$job observations, with a mean application rate of 1 percent. In the main paper, we present results using the above choice set, however, our findings are robust to using alternate cut-offs for choice set construction.[16]

3.2 Empirical Strategy

Using the above choice set, we measure the candidate distaste for distant jobs in non-metro locations by using the below specification:

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

where $Applied_{ijkt}$ 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 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.[17] The reference group includes jobs located within a candidate's home city, i.e., jobs located at zero distance from the candidate. Our first coefficient of interest β_k measures the difference in application likelihood with the distance between location i and j. If application likelihood decreases with distance, then the coefficients β_k would become more negative for higher values of k. $\mathbb{NM}_j$ is a dummy variable equal to one for jobs in non-metro, and zero otherwise, and β'_k is the coefficient of interest on the interaction term $DistGroup_k \times \mathbb{NM}_j$. A negative β'_k would suggest that candidates are less likely to apply to

[16] In general, constructing consideration sets, or as in our case arriving at the set of jobs considered by the candidate before making an application, is a pertinent research question 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) use networks of job seekers who make common job applications to arrive at the set. 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 us to account for differential effect of job and candidate attributes, as we discuss later.

[17] We use geodist command in Stata to calculate the distance between two locations. Specifications with more disaggregate groups give similar results as coefficients beyond 100 miles have similar magnitude. Our results are robust to using alternate distance groups.

jobs in a non-metro city vs. metro city, when the two jobs are equidistant.

We account for variation coming from job-level controls ($\mathbb{X}_j$) like city, qualification, experience as well as candidate-level controls ($\mathbb{X}_i$) like gender, education and age. The specification in Equation 1 also includes candidate (δ_i), job (δ_j), and month-year (δ_t) fixed-effects. δ_i filter out any unobserved candidate-level factors, δ_j control for any unobserved job-level characteristics, and δ_t account 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. Standard errors are clustered at the candidate level.

3.3 Distaste for Non-Metro Jobs

Table 1 reports the results. As first pass, we look at the impact of distance on application by excluding the non-metro interaction term in the regression. This exercise is similar to Banfi *et al.* (2019) who report that amongst young unemployed job seekers in Chile, the application rate declines by 40 percent for jobs located more than 200 miles away. Marinescu & Rathelot (2018) find that the probability of application falls to almost zero beyond 75 miles in the U.S. Columns (1)–(3) estimate the distaste for distance using a specification that represents a significant advancement over these previous studies, by controlling for job and candidate level unobservables.

Column (1) includes candidate-level controls and job fixed-effects, thereby mitigating the effect of unobserved job-level heterogeneity. Column (2) includes job-level controls with candidate fixed-effects, accounting for unobserved candidate-level variance, while column (3) 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. Under all three specifications, considering the baseline application probability within the candidate's home city is 6.5 percent, these coefficients suggest a near-zero probability of application for jobs more than 100 miles away. The magnitude of distaste among the Indian job seekers is intermediate when compared to the above-mentioned papers in the literature.[18]

Columns (4)–(6) report the differential distance-based distaste across jobs in metro vs. non-metro cities, accounting for various fixed-effects and controls in the regression. All specifications give similar quantitative results. Column (6) reports the most saturated specification, incorporating job fixed-effects, candidate fixed-effects and month-year time fixed-effects. 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 points for non-metro jobs located 500 miles away from the candidate. As a result, the average distaste for non-metro jobs is three times the distaste for a metro city job located 500 miles away.

While jobs located in distant metro cities witness distaste, the magnitude is notably smaller compared to non-metro cities. Two key observations highlight the distinct nature of distaste for non-metro jobs. First, metro jobs still have a non-zero probability of receiving applications from candidates located beyond 500 miles, which is not the case for non-metro jobs. Thus, the extreme distaste for distance is largely explained by distaste for non-metro jobs. Second, the differential distaste for non-metro jobs is 7.3 percentage points beyond 50 miles, leading to an application probability of zero beyond that threshold. This suggests that non-metro locations primarily attract applications from candidates within a 50-mile radius.

We next examine whether higher wages can offset the observed distaste for non-metro jobs (Table 2). Column (1) is the same as column (6) in Table 1.[19] In column (2), we estimate

[18] We also examine the role played by other potential factors discussed in the literature such as job ad level characteristics like candidate's gender, education and experience in influencing the candidate's distaste for distant jobs in Appendix Section B.2. While these factors are statistically significant, they, however, do not quantitatively account for the candidate's observed disinclination to apply for jobs that are located far away as the main coefficients (β_k) remain largely similar in these regressions.

[19] Around 56% of the job ads report wages in our data and we construct the potential choice set over

the distaste across metro vs. non-metro cities for jobs with posted wages. The application rate in candidate's home city is higher at 10.8% for these jobs, leading to larger coefficient estimates for distaste. We find that candidates are 7.5 percentage point or 70 percent less likely to apply to jobs in metro cities located 500 miles away. This distaste nearly triples for non-metro locations, with the probability of application dropping to zero, consistent with the findings in full sample.

It is plausible that the high distaste for non-metro jobs is on account of lower wages in non-metro jobs rather than a preference for amenities in metro cities. To test this hypothesis, we extend the specification in column (2) by including an additional control that interacts the distance groups with an indicator variable for the posted wage being above the median wage (column (3)). While above-median wage reduces distaste for distant locations, the higher distaste for non-metro jobs continues to persist even in this specification.[20]

Lastly, in column (4), we include triple interaction terms (distance$\times$above median wage$\times$non-metro). We find that the compensating effect of wages is more pronounced for non-metro jobs as the coefficients on distance$\times$above median wage are quantitatively small relative to those on triple interaction terms. Our estimates suggest that candidates would require $(18.4/6.9 = 2.66)$ times the median wage to make them indifferent between a metro and a non-metro job located 500 miles away. To make them indifferent between a job in their home city and a non-metro job more than 500 miles away, they would require $((18.4+7.8)/6.9 = 3.79)$ times the median wage.

these jobs for each candidate. This yields 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.

[20] The results are robust to using continuous wage measure (Appendix Table A.4) instead of binary wage variable, above- or below-median. We also estimate a specification with wages adjusted for cost of living (Appendix Table A.5) and get similar results.

3.4 Robustness

The above results remain robust to alternative candidate samples, additional controls, different time periods, and various data construction choices.

Candidate Ability: Appendix Table A.6 reports a higher distaste for jobs in distant non-metro locations even for the sub-sample of college-educated job seekers who in general have a lower distaste for distant jobs (Amior, 2024). Similarly, the distaste for distant non-metro candidates remains similar even after controlling for 12th-grade academic performance (Appendix Table A.7).

Other Candidate/Job-specific Controls: Our results are robust to including other candidate or job-specific controls in Appendix Table A.8. Specifically, column (4) reports the results after interacting job-specific characteristics (role, title, firm) with distance indicators. We find that higher distaste for distant non-metro locations holds even within the jobs posted by the same firm in the same job title, showing that the effect is not driven by other job characteristics that can vary across bigger vs. smaller cities.

Alternate Periods: We get similar results using data from 2019 (Appendix Table A.9). Therefore, our main results are not driven by the COVID-19 pandemic in 2020.

City Population Size: Our results remain consistent when using log(population) as a measure instead of the binary metro/non-metro classification. We continue to find that distant cities with smaller population witness higher distaste (Appendix Table A.10).

Alternate Distance Bins/City Classification/Choice Set: Our results are robust to using alternate bins for distance groups, using different number of metro cities, or using other date cutoffs for the choice set construction. These results are available on request.

4 Non-metro: Which Factors Explain Distaste?

This section investigates the role of long-distance connectivity costs, and job role concentration in explaining candidate distaste for non-metro locations. We then examine if employers in non-metro cities exhibit similar distaste towards distant candidates.

4.1 Long Distance Travel Connectivity

One of the factors why candidates may prefer metro cities is better transportation connectivity from metro to other locations. The effect of such connectivity on applications would be more pronounced if the job is located far away from candidate's home. We examine whether policies aimed at reducing long-distance travel costs, in terms of time and money (pecuniary or non-pecuniary costs), can impact candidate's preference to apply to non-metro jobs.[21]

We exploit the natural experiment setting provided by *UDAN* - *"Ude Desh ka Aam Naagrik"* scheme that enhanced air travel connectivity for smaller cities in India. Beginning April 2017, the scheme introduced new air routes at subsidized airfare, approximately at half the cost of road travel. Airlines bid for air routes, with the contract being awarded to the one requiring the least subsidy. Under the scheme, the airfare for a one-hour journey by a 'fixed wing aircraft' for about 500 km, was fixed at INR 2500 (USD 75).[22] Appendix Section B.3 provides additional institutional details on this scheme and data construction.

Given that the scheme reduces both travel time and cost between candidates' home cities and job locations, we hypothesize that it may change candidate application behavior to jobs in distant cities that are connected under the scheme. Using applications data from June 2018 – December 2019, we estimate if this improved connectivity impacts candidate

[21] 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.

[22] Under *UDAN* airport fees and other charges are 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. Aviation firms operating these routes are provided funds under the Viability Gap Funding to cover their losses due to low airfares.

applications to cities that received a route under the *UDAN* scheme.[23] For the empirical exercise, on the candidate side we keep cities that received an *UDAN* route and where at least some candidates are located in our data. We exclude larger cities with pre-existing airports, as the scheme would have minimal impact on their connectivity. 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 the above 54 cities and also get connected by at least one route under *UDAN*.[24] 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 100 city pairs connected (treated) under *UDAN* vs. 1,692[25] city pairs that do not get connected.

To estimate the impact of this policy, we compare the average applications received by a job in city k from candidates whose cities get connected through *UDAN* (treated routes) vs. those cities that do not get connected (control routes), before and after the opening of the routes. Specifically, we estimate the following model:

$$Y_{jkmt} = \sum_{\tau=-3,\tau\neq-1}^{\tau=3} \gamma_\tau Route_{km}^{\tau} + \delta_{km} + \delta_t + \delta_j + dist_{km} \times \delta_t + \epsilon_{jkmt} \tag{2}$$

where Y_{jkmt} is the number of applications received by job ad j in city k from candidates in city m with posting in month-year t. The main variable of interest, $Route_{km}^{\tau}$ 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

[23] We do not use the data from 2020 for two reasons. First, air traffic in India remained below the pre-pandemic levels until 2021 (World Bank) with large scale restrictions during the COVID lockdowns. Second, we could not access data from January 2020-May 2020 from the platform as they had erased it by the time we approached them for new data access post COVID-19.

[24] For our main estimation, we consider routes that were launched after June 2018 since the applications data begins from June 2018. Also, we keep those job location cities that receive at least one route under *UDAN*. This ensures that the subset of cities in our data are comparable to each other as they all receive a new airport route. This selection is inconsequential because cities not covered under the scheme would not matter as they get dropped in the final estimation, as we discuss later in the estimation strategy.

[25] Origin and destination matter for the analysis, hence ij and ji city pairs counted separately.

operational corresponds to $\tau = 0$.[26] γ_τ captures 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 γ_τ coefficients, i.e. those before the opening of routes, are significantly different from zero.

We control for city pair level fixed-effects (δ_{km}), time fixed-effects (δ_t), job level fixed-effects (δ_j) and distance time trends ($dist_{km} \times \delta_t$). δ_{km} accounts for non-time-varying heterogeneity for city pair (k, m), while $dist_{km} \times \delta_t$ accounts for time-varying trends at a given distance.[27] δ_j controls for any job-specific variation. Since the routes open in a staggered fashion, we use the estimation methods from the recent 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). Given these extensive controls, the staggered DID method proposed by Sun & Abraham (2021) that allows for adding covariates is best suited in our case. Standard errors are clustered at job and candidate city level (k, m) given that the treatment is at the level of city pair.[28]

We estimate Equation 2 separately for jobs located in metro vs. non-metro cities. First, it allows us to test if connectivity matters for explaining distaste for both metro and non-metro locations. Second, given metro and non-metro are significantly different from each other, it

[26]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 more than 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 (see Schmidheiny & Siegloch (2019)).

[27]The term $dist_{km} \times \delta_t$ accounts for potential 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 partially mitigated as all cities in the sample get at least one route. However, when city k gets connected to cite m_1 and not to m_2, there could be a concern that the decision to start the route between (k, m_1) is on account of unobserved factors that also influence applications over time. For instance, prior to *UDAN*, there could be a secular increase in people-to-people connection between (k, m_1), leading to more applications from k to m_1 as well as leading to opening of route (k, m_1). Additionally, we find that less distant city pairs among our sample have a higher likelihood to be connected through *UDAN*. To address this concern, we control for trends in applications across cities over time based on distance between them. However, as discussed later, results are similar without distance time trends.

[28]While we adopt the estimation method of Sun & Abraham (2021) due to the extensive set of controls in our model, Baker *et al.* (2022) note that for models with fewer controls, the estimates are numerically equivalent to those obtained using Callaway & Sant'Anna (2021).

ensures that the control routes consists of comparable job cities.

Figure 4 plots the coefficients from the event study estimates. Panel (a) reports the impact on jobs located in metro cities, while Panel (b) plots it for jobs in non-metro cities with candidates based in other non-metro locations. When a metro is connected by a *UDAN* route, there is a significant increase in job applications to the connected metro, with an increase of 0.02 applications per job. Given that the average application rate in metro cities is 0.08, this represents a 25 percent rise in applications for jobs located in a distant metro. Concurrently, no positive pre-trends are visible prior to the launch of these routes. In contrast, Panel (b) reveals no significant effect on applications from non-metro cities to other non-metro cities, even when they are connected via *UDAN*. This suggests that improved connectivity does not lead to increased application rates for jobs in non-metro cities.

Robustness and Other Tests: We estimate the above specification without distance time trends and get similar results (Appendix Figure A.1). Furthermore, we do two additional checks. First, we test for spillover effects by including cities within 50 miles of the baseline 54 non-metro cities. We find no evidence of spillover and no additional application from these cities to non-metro jobs. Second, we estimate the effect on applications to non-metro jobs by candidates located in metro cities after the launch of *UDAN* route. Again, we find no effect of the scheme. These results are available on request.

In summary, these findings underscore that the decline in connectivity costs increases job applications to metro cities. If a city receives an air route to a metro, it increases the likelihood of candidates applying to jobs located in the linked metro. However, similar connectivity to non-metro cities does not lead to a higher application rate for jobs located there. Consequently, even if non-metro locations are connected through subsidized air travel, reduced transport costs are unlikely to attract candidates, as other barriers may play a more substantial role in determining application behavior. We explore these factors below.

4.2 Concentration of Job Roles in Cities

Another important factor influencing job seekers' decisions is the concentration of specific occupations or job roles in a given city. On average, metro cities have more job opportunities, which can incentivize candidates to prefer these locations due to the perceived option value associated with better match in future job search (Glaeser & Gottlieb, 2009; Freedman, 2008; Almeida & Kogut, 1999). For instance, if finance-related jobs are predominantly located in Mumbai, candidates planning a career in this field may choose Mumbai considering better future opportunities. The possibility of being able to switch jobs within the same city is also likely to be higher when similar roles exist nearby. While larger cities, on average, have dense labor markets, occupation/role-based concentration can also exist in non-metro locations, thus reducing the distaste for distance for particular job roles in a given non-metro. For instance, the city of Hyderabad is a known pharmaceutical cluster in India and may attract individuals looking for a career in the pharma industry.

Using the job role information from our dataset, we compute the *proportion* of jobs in a specific job role posted in a city to examine the impact of job role concentration on applications. We calculate dynamic proportion of jobs posted in a city for a given job role at the candidate-level, based on the proportion of job ads for that role within the choice set of a given candidate. We use candidate's choice set, as outlined in Section 3, i.e., job roles posted between $d - 10$ and $d + 1$, to calculate the job role proportion across cities at candidate level.[29] This approach exploits the idea that candidates may update beliefs about job role concentration in a city based on the job ads they see on the portal during their job search spell. Since this candidate-level proportion variable varies over time for a job role-city pair, our identification comes from comparing application likelihood based on the dynamic value of concentration of a job role in a given city over time based on the candidate's search spell. Notably, the timing of job search for a candidate is likely to be quasi-random, allowing for

[29] The mean (s.d.) of this candidate level proportion variable is 0.15(0.14), 0.19(0.15), 0.08(0.07), for all, metro, and non-metro respectively.

causal interpretation of results. This also addresses any concerns that may arise if we instead use static average value for the proportion of jobs in a given role and city over the entire time period as it may be correlated with other non time-varying job role-city pair attributes.

We incorporate the interaction of proportion variable with distance in Equation 1 to examine the impact of job role concentration on applications. The results are reported in Table 3. Column (1) examines the effect of job role proportion as a function of distance on candidate applications across all cities, after controlling for job, candidate, and time fixed effects We find that a 10 percentage point increase in the proportion of jobs in a specific role in a city increases the likelihood of a candidate applying to a job in that city (over 500 miles away) by 1.6 percentage points.

Columns (2)–(3) report the estimates for metro and non-metro cities respectively. Analyzing the sub-samples allows us to assess the differential effect of job role concentration between these two city types. We find that a higher concentration of job roles increases the application probability for both metro and non-metro cities. However, the impact on non-metro cities is quantitatively more significant. Specifically, a 10 percentage point increase in the proportion of jobs within a given job role-city pair makes candidates 0.7 percentage points more likely to apply to a metro location and 8 percentage points more likely to apply to a non-metro location, both located 500 miles away. In relative terms, for the same distance, a 10 percentage point increase in proportion reduces distaste for metro by $(0.07/6 = 0.116)$ 11 percent. In contrast, for non-metro, the same change results in reducing the distaste by a more substantial $(0.08/0.21 = 0.38)$ 38 percent. These results suggest that non-metro locations with more concentrated job roles are able to mitigate a significant portion of the distaste associated with non-metro.

Robustness: The above results are robust to additional controls (Appendix Table A.11). Columns (1) and (2) controls for city×job role×distance fixed effects, akin to controlling for static proportion. Columns (3) and (4) control for similar job role in a given firm. Columns (5) and (6) provide the results after controlling for the interaction of job title, firm, and distance, i.e., comparing similar positions within a given firm at similar distance from candidates. We also find similar results if we instead use static proportion at city-job role level (Appendix Table A.12). Relative to the dynamic concentration measure based of candidate search spell in Table 3, here we compute concentration for the entire sample period.

Role of Other City-level Attributes: In Appendix Table A.13, we account for the differential effect of other factors that could potentially explain distaste for distant non-metro jobs. It is possible that the coefficients on distance×non-metro are picking up the effect of other city attributes that are more correlated with non-metro locations. While using dynamic candidate-specific concentration measure minimizes this concern, we still conduct additional robustness checks. Specifically, we rule out social connectivity and language proximity between city pairs as potential confounding variables. Non-metro cities may have less cultural proximity to a candidate's home city, as they tend to be less diverse compared to metro cities, which could result in fewer applications for non-metro jobs. Columns (1) and (2) accounts for social connectivity using Social Connectivity Index (SCI) as in Bailey *et al.* (2018). Even after accounting for SCI, we continue to find that job role concentration plays a crucial role for non-metro application.[30] Similarly, columns (3) and (4) provide consistent results even after controlling for distance×language proximity.[31]

[30]There is a large literature studying the role of networks and international migration (Borjas, 1992; Mayda, 2010; Munshi, 2014; McKenzie & Rapoport, 2007). Some recent work has looked at the role of networks in within country migration (Basu *et al.*, 2022; Bailey *et al.*, 2018). Migrants can be concentrated in a specific job role or cities, giving rise to immigrant niches (see survey paper by Eckstein & Peri (2018)).

[31]Language proximity measures the language overlap between the job and candidate location state similar to the Linguistic distance index as in Fenske & Kala (2021). Our measure includes only the same language overlap between the two states – Linguistic Distance$_{o,d} = \sum_{m=1}^{23} (s_{o,m} \times s_{d,m})$ where $s_{o,m}$ is the share of speakers of language m in state o, $s_{d,m}$ is the share of speakers of language m in state d. Using the Census 2011 data, we calculate the proportion of each of the 23 languages in each state to calculate this measure.

4.3 Connecting Cities with High Job Role Concentration

The above findings underscore the significant influence of job role concentration on candidate applications to non-metro jobs, while transport connectivity appears to have no effect. We further investigate whether transport connectivity has an impact on applications when a non-metro has a high job role concentration. Specifically, we estimate equation 2 separately for cases with job role proportions above or below the median.

These results are reported in Figure 5. Panel (a) displays the results for non-metro locations with below-median job role concentration, where we find that the introduction of a new route does not impact the number of applications per job. However, in Panel (b), we observe that introducing a route increases applications to non-metro locations when the city has above-median concentration in a given job role. This pattern mirrors the effect seen in metro cities with above-median concentration in Panel (c) as well as the baseline results for metro cities reported in Figure 4.[32]

Juxtaposing these findings 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. Transportation connectivity may further enhance this appeal, building upon the pre-existing attractiveness of non-metro for job seekers. However, transportation connectivity alone is insufficient to make non-metro locations attractive for job applicants.

4.4 Employer Distaste for Distant Candidates

Our analyses thus far has analyzed the candidate distaste for distant non-metro jobs. However, such distaste could be an equilibrium outcome where employers exhibit a similar distaste for candidates residing far away, thereby making candidates' choice to not apply to distant non-metro jobs optimal. Using shortlisting information, we examine if employers demonstrate

[32]We do not report the below median proportion case for metro as there are few observations with metro jobs having below median concentration, preventing consistent estimation. Instead, we can perform the same analyses with above/below 75th percentile, and find similar results as non-metro. Specifically, job roles with above the 75th percentile proportion see an increase in applications in both metro and non-metro jobs. These results are available upon request.

any such distaste. We estimate the following equation:

$$Shortlist_{ijkt} = \theta_0 + \sum_{k=1}^{g} \theta_k DistGroup_k + \sum_{k=1}^{g} \theta'_k DistGroup_k \times \mathbb{NM}_j + \lambda_i \mathbb{X}_i + \lambda_j \mathbb{X}_j + \mu_i + \mu_j + \mu_t + \varepsilon_{ijkt} \tag{3}$$

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 θ_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. Similarly, a negative θ'_k shows that firms in non-metro are less likely to shortlist distant candidates.

We control for unobservable factors at the candidate (μ_i) and job (μ_j) level. 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 are not 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 job ad to give equal weight to all job ads in our regression. The standard errors are clustered at the job level.

Table 4 reports these results. We first document the average distaste for distant candidates for all job locations (Column (1)). The decline in shortlisting probability based on distance is modest, standing at 4.9 percentage points (14 percent), for candidates located more than 500 miles away. No distaste is seen for candidates located upto 500 miles. Comparing the findings for employer distaste with candidate distaste reveals two key insights. First, candidate distaste for distance rises exponentially for jobs located beyond 100 miles, whereas employer distaste is non-existent until candidates are more than 500 miles away. Second, the candidate's 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. These findings suggest that the distaste for distance observed among job seekers is unlikely to be an equilibrium outcome driven by employer preferences for local candidates.[33]

Next, we examine differential distaste for distant candidates by metro vs. non-metro employers. Columns (2)–(4) report the heterogeneous estimates using a battery of fixed-effect and controls. 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 20 percent of the outcome mean within the city. Conversely, employers in non-metro are 11-17 percentage points more likely to shortlist candidates located 500 miles away than employers in metro cities (column (4)). This positive effect in the case of non-metro jobs nullifies the baseline distaste for candidates beyond 500 miles.[34] This table shows that candidate distaste for non-metro jobs is not reciprocated by a similar distaste on the employer side. In fact, if anything, employers in non-metro cities are more likely to shortlist candidates from distant locations compared to metro employers.

The non-discrimination for distant candidates by non-metro employers can be explained by the fact that non-metro jobs receive fewer applications on average. We provide evidence for this in Table A.16 with applications per vacancy in a job ad as our outcome variable. Columns (1) and (2) are based on the sample of job ads for which we have shortlisting data, while columns (3)–(5) are based on all job ads. We find that the coefficient on non-metro is always negative and significant, showing that jobs in non-metro locations always receive significantly lower number of applications. In terms of magnitude, among the jobs for which we observe shortlisting, non-metro cities receive lower applications by almost 55 per vacancy. For all jobs this comes down to 9 applications per vacancy. In general, non-metro locations

[33] The observed employer distaste for candidates located over 500 miles away is partially attributed to a lower likelihood of shortlisting female candidates from distant locations. This may stem from an assumption that female candidates from distant cities are less likely to accept an offer if selected (Appendix Table A.14).

[34] We report the robustness of our results to including jobs with job roles as 'others' in Appendix Table A.15. All results continue to hold.

receive an almost 33% lower average number of applications per vacancy. Given the smaller candidate pool available, this explains why employers in non-metro locations do not exhibit distaste for distant candidates.

5 Discussion and Conclusion

Our paper presents evidence for divergence in spatial search behavior of job seekers depending on whether the job is located in a metro vs. non-metro location. In the context of India, our findings align with the migration or urban trends identified in the broader spatial development literature (Colmer, 2015; Desmet *et al.*, 2015), with rising spatial inequality and high-density clusters explaining majority of economic growth in the country. While big cities continue to attract workers, often leading to congestion, smaller cities fail to attract workers.

Our findings have significant implications for policies aimed at incentivizing economic activity in targeted areas outside the major metro cities. Much like the European Union (Ehrlich & Overman, 2020) and US (Bartik, 2020), India too has placed strong emphasis on regional economic development through various policies like the Smart Cities Mission, or the *UDAN* scheme as analyzed here. Our results indicate that measures such as improving connectivity of non-metro cities or incentivizing firms in these locations to recruit distant candidates may be ineffective in attracting workers to non-metro locations. Only increasing job role concentration may be a viable policy tool for making non-metro locations attractive to workers. Dense labor markets in specialized roles in small cities mitigate career risks, making them more attractive to job seekers. In essence, our analyses answer the broader question – is it easier to move people or to move jobs? Our findings suggest that moving jobs, to create dense labor markets, is the more effective strategy. Once a significant number of jobs in a given occupation and city are available, workers will follow.

References

Abeberese, Ama Baafra, Chaurey, Ritam, & Menon, Radhika. 2024. Place-based policies and migration: Evidence from india. *The World Bank Economic Review*, lhae052.

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

Almeida, Paul, & Kogut, Bruce. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management science*, **45**(7), 905–917.

Amior, Michael. 2024. Education and geographical mobility: The role of the job surplus. *American Economic Journal: Economic Policy*, **16**(4), 341–381.

Arntz, Melanie, Brüll, Eduard, & Lipowski, Cäcilia. 2023. Do preferences for urban amenities differ by skill? *Journal of Economic Geography*, **23**(3), 541–576.

Atkins, Rachel MB, Hernández-Lagos, Pablo, Jara-Figueroa, Cristian, & Seamans, Robert. 2023. JUE Insight: What is the impact of opportunity zones on job postings? *Journal of Urban Economics*, **136**, 103545.

Bailey, Michael, Cao, Rachel, Kuchler, Theresa, Stroebel, Johannes, & Wong, Arlene. 2018. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, **32**(3), 259–280.

Baker, Andrew C, Larcker, David F, & Wang, Charles CY. 2022. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, **144**(2), 370–395.

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*.

Bartik, Timothy J. 2020. Using place-based jobs policies to help distressed communities. *Journal of Economic Perspectives*, **34**(3), 99–127.

Basu, Arnab K, Chau, Nancy H, & Lin, Gary. 2022. Migration Gravity, Networks, and Unemployment.

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

Baum-Snow, Nathaniel, & Pavan, Ronni. 2012. Understanding the city size wage gap. *The Review of economic studies*, **79**(1), 88–127.

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.

Bondonio, Daniele, & Greenbaum, Robert T. 2007. Do local tax incentives affect economic growth? What mean impacts miss in the analysis of enterprise zone policies. *Regional science and urban economics*, **37**(1), 121–136.

Borjas, George J. 1992. Ethnic capital and intergenerational mobility. *The Quarterly journal of economics*, **107**(1), 123–150.

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.

Bryan, Gharad, & Morten, Melanie. 2015. Economic development and the spatial allocation of labor: Evidence from indonesia. *Manuscript, London School of Economics and Stanford University*, 1671–1748.

Bryan, Gharad, Glaeser, Edward, & Tsivanidis, Nick. 2020. Cities in the developing world. *Annual Review of Economics*, **12**, 273–297.

Busso, Matias, Gregory, Jesse, & Kline, Patrick. 2013. Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, **103**(2), 897–947.

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

Chaturvedi, Sugat, Mahajan, Kanika, & Siddique, Zahra. 2024. Using Domain-Specific Word Embeddings to Examine the Demand for Skills. *Pages 171–223 of: Big Data Applications in Labor Economics, Part B*. Emerald Publishing Limited.

Chaurey, Ritam. 2017. Location-based tax incentives: Evidence from India. *Journal of Public Economics*, **156**, 101–120.

Chaurey, Ritam, Mahajan, Kanika, & Tomar, Shekhar. 2024. *Trumping Immigration: Visa Uncertainty and Jobs Relocation*. Tech. rept. Available at SSRN 4753372.

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.

Colmer, Jonathan. 2015. Urbanisation, growth, and development: evidence from India. *Working paper*.

Criscuolo, Chiara, Martin, Ralf, Overman, Henry G, & Van Reenen, John. 2019. Some causal effects of an industrial policy. *American Economic Review*, **109**(1), 48–85.

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

De la Roca, Jorge, Ottaviano, Gianmarco IP, & Puga, Diego. 2023a. City of dreams. *Journal of the European Economic Association*, **21**(2), 690–726.

De la Roca, Jorge, Parkhomenko, Andrii, & Velásquez-Cabrera, Daniel. 2023b. *Skill Allocation and Urban Amenities in the Developing World*. Tech. rept. Working paper.

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.

Eckstein, Susan, & Peri, Giovanni. 2018. Immigrant niches and immigrant networks in the US labor market. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, **4**(1), 1–17.

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

Ehrlich, Maximilian v, & Overman, Henry G. 2020. Place-based policies and spatial disparities across European cities. *Journal of economic perspectives*, **34**(3), 128–149.

Falck, Oliver, Heblich, Stephan, & Kipar, Stefan. 2010. Industrial innovation: Direct evidence from a cluster-oriented policy. *Regional Science and Urban Economics*, **40**(6), 574–582.

Fenske, James, & Kala, Namrata. 2021. Linguistic Distance and Market Integration in India. *The Journal of Economic History*, **81**(1), 1–39.

Freedman, Matthew, Khanna, Shantanu, & Neumark, David. 2023. Jue insight: The impacts of opportunity zones on zone residents. *Journal of Urban Economics*, **133**, 103407.

Freedman, Matthew L. 2008. Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. *Journal of Urban Economics*, **64**(3), 590–600.

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

Gibbons, Stephen, Lyytikäinen, Teemu, Overman, Henry G, & Sanchis-Guarner, Rosa. 2019. New road infrastructure: the effects on firms. *Journal of Urban Economics*, **110**, 35–50.

Glaeser, Edward L, & Gottlieb, Joshua D. 2009. The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of economic literature*, **47**(4), 983–1028.

Glaeser, Edward L, Kolko, Jed, & Saiz, Albert. 2001. Consumer city. *Journal of economic geography*, **1**(1), 27–50.

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

Grogger, Jeffrey, & Hanson, Gordon H. 2011. Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, **95**(1), 42–57.

Ham, John C, Swenson, Charles, İmrohoroğlu, Ayşe, & Song, Heonjae. 2011. Government programs can improve local labor markets: Evidence from state enterprise zones, federal empowerment zones and federal enterprise community. *Journal of Public Economics*, **95**(7-8), 779–797.

Hasan, Rana, Jiang, Yi, & Rafols, Radine Michelle. 2021. Place-based preferential tax policy and industrial development: Evidence from Indiaâs program on industrially backward districts. *Journal of Development Economics*, **150**, 102621.

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

Huber, Franz. 2012. Do clusters really matter for innovation practices in Information Technology? Questioning the significance of technological knowledge spillovers. *Journal of economic geography*, **12**(1), 107–126.

Hyman, Benjamin G, Freedman, Matthew, Khanna, Shantanu, & Neumark, David. 2022. *Firm responses to state hiring subsidies: Regression discontinuity evidence from a tax credit formula*. Tech. rept. National Bureau of Economic Research.

Kline, Patrick, & Moretti, Enrico. 2013. Place based policies with unemployment. *American Economic Review*, **103**(3), 238–243.

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

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.

Lu, Yi, Wang, Jin, & Zhu, Lianming. 2019. Place-based policies, creation, and agglomeration economies: Evidence from Chinaâs economic zone program. *American Economic Journal: Economic Policy*, **11**(3), 325–360.

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.

Martin, Philippe, Mayer, Thierry, & Mayneris, Florian. 2011. Public support to clusters: A firm level study of French âLocal Productive Systemsâ. *Regional Science and Urban Economics*, **41**(2), 108–123.

Marx, Benjamin, Stoker, Thomas, & Suri, Tavneet. 2013. The economics of slums in the developing world. *Journal of Economic perspectives*, **27**(4), 187–210.

Mayda, Anna Maria. 2010. International migration: A panel data analysis of the determinants of bilateral flows. *Journal of population economics*, **23**, 1249–1274.

Mayer, Thierry, Mayneris, Florian, & Py, Loriane. 2017. The impact of Urban Enterprise Zones on establishment location decisions and labor market outcomes: evidence from France. *Journal of Economic Geography*, **17**(4), 709–752.

McDonald, Frank, Huang, Qihai, Tsagdis, Dimitrios, & Josef Tüselmann, Heinz. 2007. Is there evidence to support Porter-type cluster policies? *Regional studies*, **41**(1), 39–49.

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

McKenzie, David, & Rapoport, Hillel. 2007. Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of development Economics*, **84**(1), 1–24.

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.

Moretti, Enrico. 2014. Cities and growth. *International Growth Center.*

Munshi, Kaivan. 2014. Community networks and the process of development. *Journal of Economic Perspectives*, **28**(4), 49–76.

Neumark, David, & Kolko, Jed. 2010. Do enterprise zones create jobs? Evidence from Californiaâs enterprise zone program. *Journal of Urban Economics*, **68**(1), 1–19.

Neumark, David, & Simpson, Helen. 2015. Place-based policies. *Pages 1197–1287 of: Handbook of regional and urban economics*, vol. 5. Elsevier.

Porcher, Charly. 2020. Migration with Costly Information.

Redding, Stephen J, & Turner, Matthew A. 2015. Transportation costs and the spatial organization of economic activity. *Handbook of regional and urban economics*, **5**, 1339–1398.

Roca, Jorge De La, & Puga, Diego. 2017. Learning by working in big cities. *The Review of Economic Studies*, **84**(1), 106–142.

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.

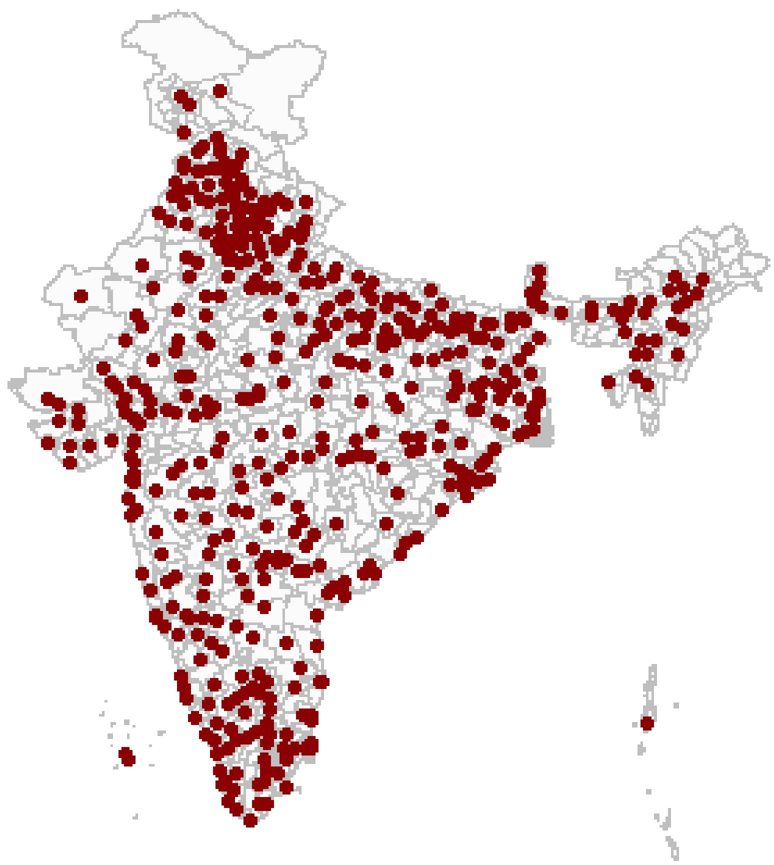
Sun, Liyang, & Abraham, Sarah. 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, **225**(2), 175–199.

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.

Wang, Jin. 2013. The economic impact of special economic zones: Evidence from Chinese municipalities. *Journal of development economics*, **101**, 133–147.

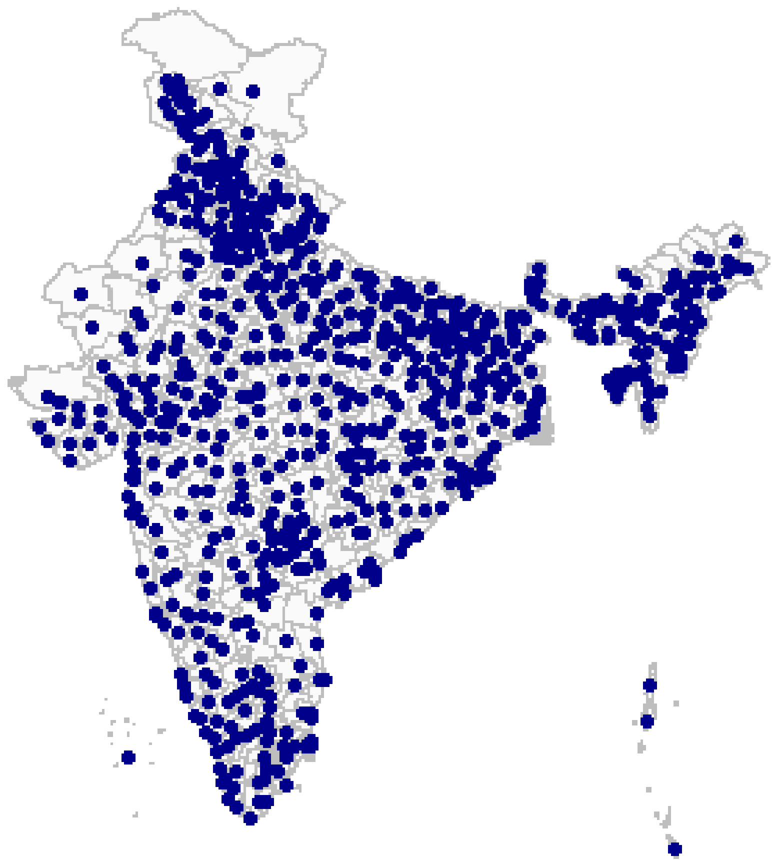
Wozniak, Abigail. 2010. Are college graduates more responsive to distant labor market opportunities? *Journal of Human Resources*, **45**(4), 944–970.

Figure 1: Location of jobs



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

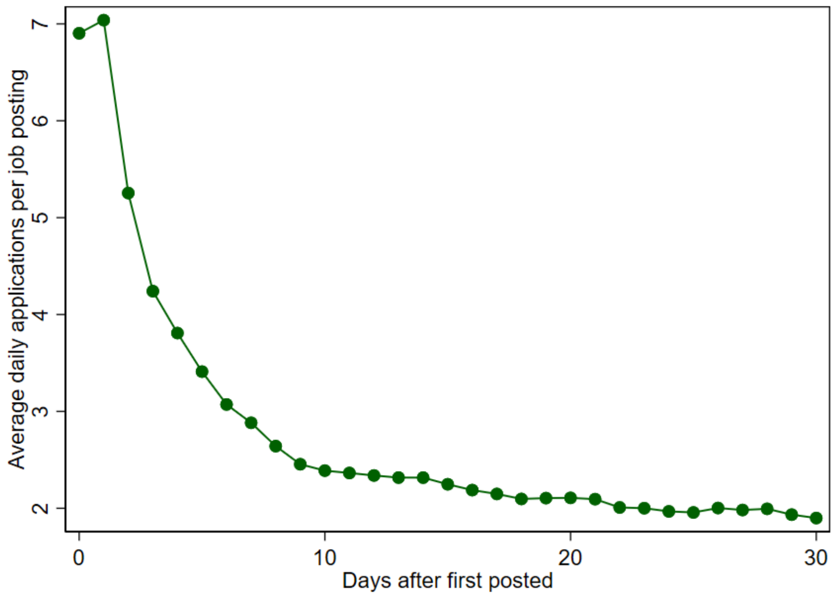
Figure 2: Location of candidates



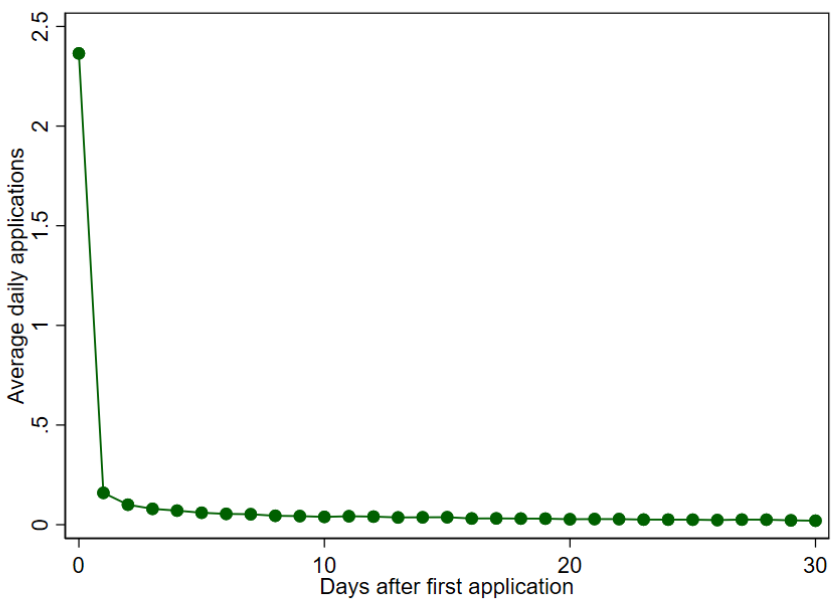
Notes: The map plots the locations for the candidates who apply to the posted jobs over the period July - December 2020.

Figure 3: Activity on Posted Jobs

(a) Average Applications from Date of Job Posting



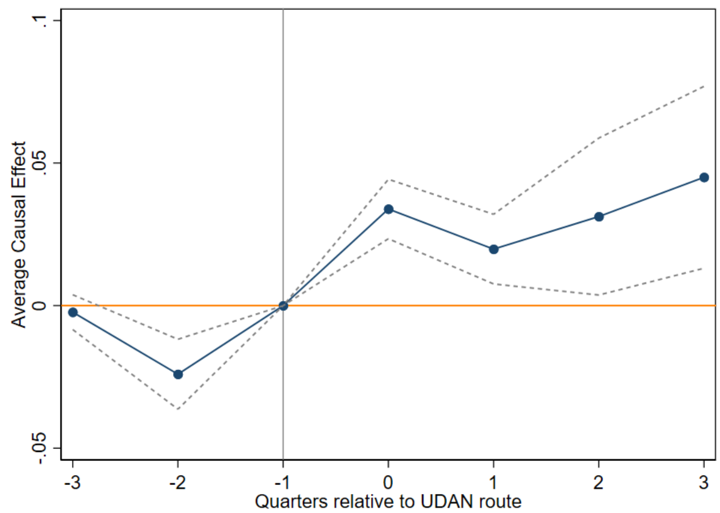
(b) Average Daily Applications by Candidates after First Application



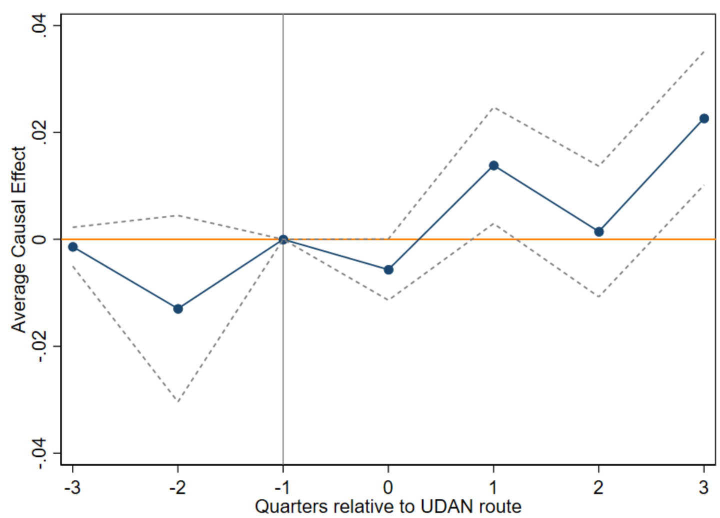
Notes: Panel (a) shows the daily average of applications received on a job since it's posting date for the period July-December 2020. Jobs receive maximum applications immediately after the job posting and it falls almost flat beyond the 10 days period. Here, Day0 is defined as job posting date. Same 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 4: *UDAN*: Effect of Increased Connectivity on Job Applications with distance time trends

(a) Metro jobs



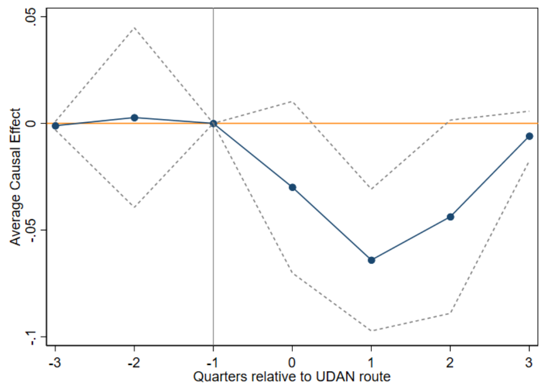
(b) Non-metro jobs



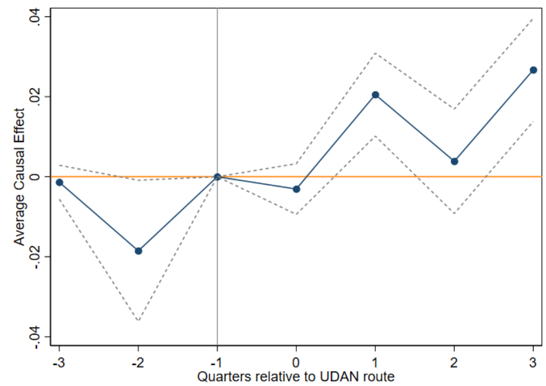
Notes: We plot the impact of air routes becoming operational between a city pair on applications received for the posted job. The outcome of interest is the number of applications received by a job in city k from candidates in city m. We have excluded applications for job role 'Others'. Event-study plots 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. Co-variates include distance and time trends.Standard errors are clustered at city pair level. Panel (a) reports the impact on jobs located in metro cities. When a metro is connected by a UDAN route, there is a significant increase in job applications to the connected metro, with an increase of 0.02 applications per job. Concurrently, no positive pre-trends are visible prior to the launch of these routes. Panel (b) reveals no significant effect on applications from non-metro cities to other non-metro cities, even when they are connected via UDAN.

Figure 5: *UDAN*: Effect of Increased Connectivity on Job Applications as a Function of Job Role Concentration

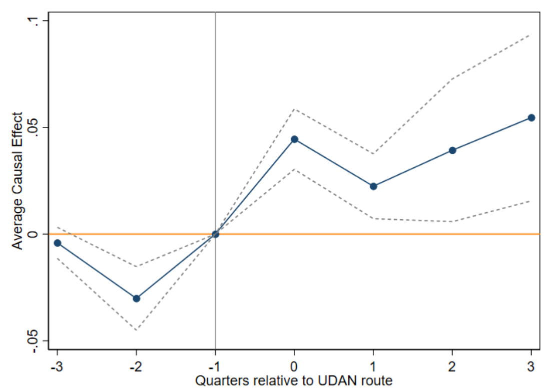
(a) Non-metro Jobs: Below Median Concentration



(b) Non-metro Jobs: Above Median Concentration



(c) Metro Jobs: Above Median Concentration



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). All specifications include city pair, posting date and job ad fixed effects. Co-variates include distance and time trends. Panel (a) displays the results for non-metro locations with below-median job role concentration, where we find that the introduction of a new route does not impact the number of applications per job. In Panel (b), we observe that introducing a route increases applications to non-metro locations when the city has above-median concentration in a given job role. In Panel (c) observe that introducing a route increases applications to metro cities with above-median concentration. We do not report the below median proportion case for metro as there are few observations with metro jobs having below median concentration, preventing consistent estimation.

Table 1: Effect of Distance on Application by Candidates

	(1)	(2)	(3)	(4)	(5)	(6)
1-50	–0.015***	–0.017***	–0.015***	–0.013***	–0.016***	–0.013***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
50-100	–0.048***	–0.053***	–0.051***	–0.030***	–0.034***	–0.032***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
100-500	–0.056***	–0.062***	–0.059***	–0.039***	–0.043***	–0.041***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
> 500	–0.059***	–0.065***	–0.062***	–0.043***	–0.048***	–0.046***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
1-50 x 1(Non-metro)				–0.043***	–0.044***	–0.045***
				(0.003)	(0.004)	(0.003)
50-100 x 1(Non-metro)				–0.073***	–0.075***	–0.073***
				(0.001)	(0.001)	(0.001)
100-500 x 1(Non-metro)				–0.074***	–0.076***	–0.073***
				(0.001)	(0.001)	(0.001)
> 500 x 1(Non-metro)				–0.075***	–0.077***	–0.074***
				(0.001)	(0.001)	(0.001)
N	164769978	166487421	166555050	164769978	166487421	166555050
Mean Y	0.010	0.010	0.010	0.010	0.010	0.010
Mean Y (Same city)	0.065	0.065	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. $\mathbb{1}$(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. Job controls include education and experience required. Candidate controls include the candidate's gender, age, age square, and education. 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: Heterogeneity by Posted Wages

Sample:	All jobs	Jobs with posted wages		
	(1)	(2)	(3)	(4)
1-50	–0.013***	–0.023***	–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)
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. $\mathbb{1}$(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. $\mathbb{1}$(wage$>$ med) is a binary variable takes the value 1 if the offered wage for the job is above median. 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 Distance on Application: Heterogeneity by Job Role Concentration

	All (1)	Metro (2)	Non-metro (3)
1-50	–0.042***	–0.022***	–0.176***
	(0.002)	(0.002)	(0.008)
50-100	–0.081***	–0.055***	–0.201***
	(0.001)	(0.001)	(0.002)
100-500	–0.089***	–0.057***	–0.213***
	(0.001)	(0.001)	(0.002)
> 500	–0.092***	–0.060***	–0.217***
	(0.001)	(0.001)	(0.002)
Proportion	–0.151***	–0.067***	–0.786***
	(0.003)	(0.003)	(0.019)
1-50 x Proportion	0.127***	0.048***	1.528***
	(0.007)	(0.007)	(0.353)
50-100 x Proportion	0.169***	0.101***	0.824***
	(0.005)	(0.004)	(0.022)
100-500 x Proportion	0.175***	0.085***	0.861***
	(0.003)	(0.002)	(0.019)
>500 x Proportion	0.161***	0.072***	0.801***
	(0.003)	(0.003)	(0.019)
N	166555050	102007876	64546823
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. Proportion variable sums up to 1 across job locations for each job role for a given candidate within the search spell. 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 Shortlisting by Employer

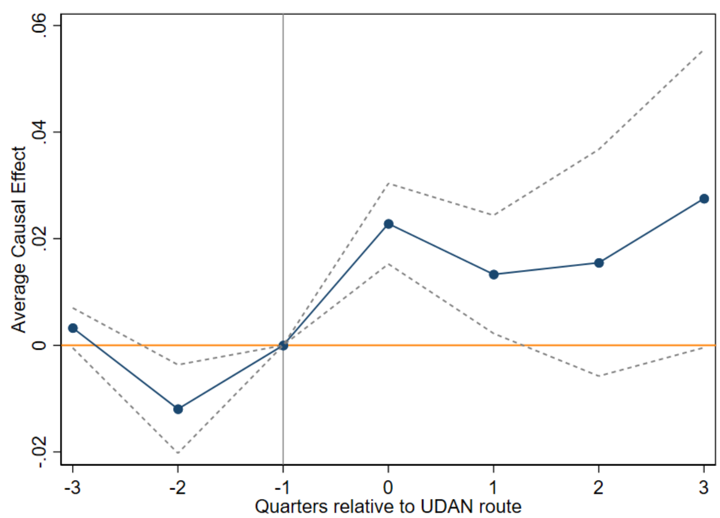
	(1)	(2)	(3)	(4)
1-50	0.010	0.080	0.149	0.147
	(0.067)	(0.048)	(0.116)	(0.119)
50-100	–0.035	0.022	–0.011	0.015
	(0.066)	(0.034)	(0.083)	(0.082)
100-500	–0.001	–0.020	–0.059**	–0.049
	(0.022)	(0.016)	(0.028)	(0.032)
> 500	–0.049**	–0.066***	–0.099***	–0.081***
	(0.020)	(0.015)	(0.025)	(0.029)
1-50 x $\mathbb{1}$(Non-metro)		–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)
N	28158	55738	28158	28158
Mean Y (Same city, base)	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. $\mathbb{1}$(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. 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.

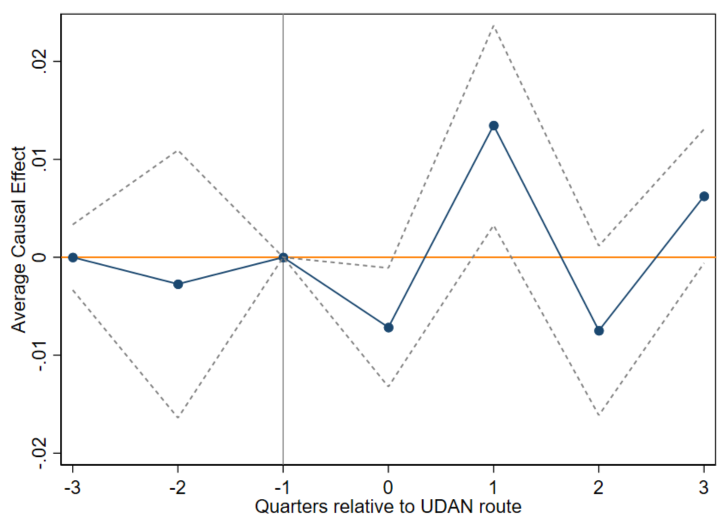
A Appendix: Additional Figures and Tables

Figure A.1: Robustness: *UDAN*: Effect of Increased Connectivity on Job Applications without 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. Standard errors are clustered at city pair level. Panel (a) reports the impact on jobs located in metro cities. When a metro is connected by a UDAN route, there is a significant increase in job applications to the connected metro, and no positive pre-trends are visible prior to the launch of these routes. Panel (b) reveals no significant effect on applications from non-metro cities to other non-metro cities, even when they are connected via UDAN.

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

Table A.4: Robustness: Effect of Distance on Applicationby candidates - Heterogeneity by Posted Wages

	(3)	(1)	(2)
1-50	–0.023***	–0.435***	–0.071
	(0.002)	(0.042)	(0.049)
50-100	–0.053***	–0.367***	–0.138***
	(0.001)	(0.016)	(0.017)
100-500	–0.066***	–0.355***	–0.153***
	(0.000)	(0.010)	(0.010)
$>$ 500	–0.075***	–0.346***	–0.160***
	(0.000)	(0.010)	(0.010)
1-50 x $\mathbb{1}$(Non-metro)	–0.111***	–0.109***	–1.298***
	(0.004)	(0.004)	(0.086)
50-100 x $\mathbb{1}$(Non-metro)	–0.147***	–0.147***	–1.039***
	(0.002)	(0.002)	(0.037)
100-500 x $\mathbb{1}$(Non-metro)	–0.152***	–0.152***	–0.992***
	(0.002)	(0.002)	(0.028)
$>$ 500 x $\mathbb{1}$(Non-metro)	–0.150***	–0.151***	–0.975***
	(0.002)	(0.002)	(0.029)
1-50 x ln (wage)		0.034***	0.004
		(0.003)	(0.004)
50-100 x ln (wage)		0.026***	0.007***
		(0.001)	(0.001)
100-500 x ln (wage)		0.023***	0.007***
		(0.001)	(0.001)
$>$ 500 x ln (wage)		0.022***	0.007***
		(0.001)	(0.001)
1-50 x $\mathbb{1}$(Non-metro) x ln (wage)			0.097***
			(0.007)
50-100 x $\mathbb{1}$(Non-metro) x ln (wage)			0.072***
			(0.003)
100-500 x $\mathbb{1}$(Non-metro) x ln (wage)			0.068***
			(0.002)
$>$500 x $\mathbb{1}$(Non-metro) x ln (wage)			0.067***
			(0.002)
N	62575521	62575521	62575521
Mean Y	0.018	0.018	0.018
Mean Y (Same city)	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. $\mathbb{1}$(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. Ln(wage) is a continuous variable, taking the natural log of the posted wage. 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 A.5: Robustness: Effect of Distance on Application - Heterogeneity by Cost of Living Adjusted Wages

Sample:	(1)	(2)	(3)
1-50	–0.023***	–0.028***	–0.027***
	(0.002)	(0.003)	(0.003)
50-100	–0.053***	–0.061***	–0.061***
	(0.001)	(0.001)	(0.001)
100-500	–0.066***	–0.073***	–0.072***
	(0.000)	(0.001)	(0.001)
> 500	–0.075***	–0.082***	–0.082***
	(0.000)	(0.001)	(0.001)
1-50 x 𝟙(Non-metro)	–0.111***	–0.111***	–0.115***
	(0.004)	(0.004)	(0.007)
50-100 x 𝟙(Non-metro)	–0.147***	–0.148***	–0.147***
	(0.002)	(0.002)	(0.003)
100-500 x 𝟙(Non-metro)	–0.152***	–0.153***	–0.155***
	(0.002)	(0.002)	(0.002)
> 500 x 𝟙(Non-metro)	–0.150***	–0.152***	–0.152***
	(0.002)	(0.002)	(0.002)
1-50 x 𝟙(wage> med)		0.013***	0.011***
		(0.004)	(0.004)
50-100 x 𝟙(wage> med)		0.019***	0.021***
		(0.001)	(0.001)
100-500 x 𝟙(wage> med)		0.016***	0.015***
		(0.001)	(0.001)
> 500 x 𝟙(wage> med)		0.016***	0.016***
		(0.001)	(0.001)
1-50 x 𝟙(wage> med) x 𝟙(Non-metro)			0.006
			(0.008)
50-100 x 𝟙(wage> med) x 𝟙(Non-metro)			–0.002
			(0.003)
100-500 x 𝟙(wage> med) x 𝟙(Non-metro)			0.003
			(0.003)
>500 x 𝟙(wage> med) x 𝟙(Non-metro)			0.001
			(0.003)
N	62575521	62575521	62575521
Mean Y	0.018	0.018	0.018
Mean Y (Same city)	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. 𝟙(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. 𝟙(wage> med) is a binary variable takes the value 1 if the offered adjusted wage is above median. Wages offered is adjusted based on cost of living for different cities from Livingcost.org as of October 2024. 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 A.6: Robustness: Effect of Distance on Application - Heterogeneity by Education Level Graduates and Above

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 $\mathbb{1}$(Non-metro)	–0.037***	–0.102***	–0.101***	–0.145***
	(0.003)	(0.005)	(0.005)	(0.006)
50-100 x $\mathbb{1}$(Non-metro)	–0.063***	–0.136***	–0.136***	–0.168***
	(0.001)	(0.002)	(0.002)	(0.003)
100-500 x $\mathbb{1}$(Non-metro)	–0.063***	–0.141***	–0.141***	–0.173***
	(0.001)	(0.002)	(0.002)	(0.002)
> 500 x $\mathbb{1}$(Non-metro)	–0.064***	–0.139***	–0.139***	–0.170***
	(0.001)	(0.002)	(0.002)	(0.002)
1-50 x $\mathbb{1}$(wage> med)			0.023***	–0.002
			(0.004)	(0.005)
50-100 x $\mathbb{1}$(wage> med)			0.017***	0.003**
			(0.001)	(0.002)
100-500 x $\mathbb{1}$(wage> med)			0.017***	0.004***
			(0.001)	(0.001)
> 500 x $\mathbb{1}$(wage> med)			0.016***	0.004***
			(0.001)	(0.001)
1-50 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.093***
				(0.009)
50-100 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.066***
				(0.004)
100-500 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.064***
				(0.003)
>500 x $\mathbb{1}$(wage> med) x $\mathbb{1}$(Non-metro)				0.063***
				(0.003)
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. $\mathbb{1}$(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. $\mathbb{1}$(wage> med) is a binary variable takes the value 1 if the offered wage for the job is above median. 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 A.7: Robustness: Effect of Distance on Application - Heterogeneity by Candidate Ability

Sample:	Graduates & above	Candidate ability	
	(1)	(2)	(3)
1-50	-0.013***	-0.012***	-0.013***
	(0.002)	(0.002)	(0.002)
50-100	-0.030***	-0.036***	-0.033***
	(0.000)	(0.001)	(0.001)
100-500	-0.038***	-0.046***	-0.044***
	(0.000)	(0.000)	(0.000)
Above 500	-0.043***	-0.050***	-0.048***
	(0.000)	(0.000)	(0.000)
1-50 x 𝟙(Non-metro)	-0.037***	-0.047***	-0.049***
	(0.003)	(0.003)	(0.004)
50-100 x 𝟙(Non-metro)	-0.063***	-0.076***	-0.085***
	(0.001)	(0.001)	(0.002)
100-500 x 𝟙(Non-metro)	-0.063***	-0.076***	-0.084***
	(0.001)	(0.001)	(0.002)
Above 500 x 𝟙(Non-metro)	-0.064***	-0.076***	-0.085***
	(0.001)	(0.001)	(0.002)
1-50 x 𝟙(Marks>75th ptile)		-0.002	0.002
		(0.003)	(0.003)
50-100 x 𝟙(Marks>75th ptile)		0.010***	0.005***
		(0.001)	(0.001)
100-500 x 𝟙(Marks>75th ptile)		0.011***	0.008***
		(0.001)	(0.001)
Above 500 x 𝟙(Marks>75th ptile)		0.010***	0.006***
		(0.001)	(0.001)
1-50 x 𝟙(Marks>75th ptile) x 𝟙(Non-metro)			0.003
			(0.006)
50-100 x 𝟙(Marks>75th ptile) x 𝟙(Non-metro)			0.019***
			(0.002)
100-500 x 𝟙(Marks>75th ptile) x 𝟙(Non-metro)			0.016***
			(0.002)
Above 500 x 𝟙(Marks>75th ptile) x 𝟙(Non-metro)			0.017***
			(0.002)
N	151665025	151061161	151061161
Controls			
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. 𝟙(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. 𝟙(Marks>75th ptile) is a binary variable that takes the value 1 if the candidate's marks in twelfth standard are above 75th percentile, takes 0 otherwise. 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 A.8: Robustness: Effect of Distance on Application by candidates - Additional controls

	(1)	(2)	(3)	(4)
1-50	–0.023***			
	(0.002)			
50-100	–0.058***			
	(0.001)			
100-500	–0.070***			
	(0.001)			
$>$ 500	–0.074***			
	(0.001)			
1-50 x $\mathbb{1}$(Non-metro)	–0.061***	–0.057***	–0.048***	–0.024***
	(0.003)	(0.002)	(0.003)	(0.004)
50-100 x $\mathbb{1}$(Non-metro)	–0.078***	–0.072***	–0.079***	–0.037***
	(0.001)	(0.001)	(0.001)	(0.001)
100-500 x $\mathbb{1}$(Non-metro)	–0.081***	–0.073***	–0.082***	–0.039***
	(0.001)	(0.001)	(0.001)	(0.001)
$>$ 500 x $\mathbb{1}$(Non-metro)	–0.084***	–0.077***	–0.085***	–0.041***
	(0.001)	(0.001)	(0.001)	(0.001)
N	166487070	166487070	166484851	166124987
Controls				
Candidate FE	✓	✓	✓	✓
Job FE	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓
Candidate FE × Non-metro	✓	✓	✓	✓
Job edu, exp × distance	✓	✓	✓	✓
Job role × distance		✓		
Job title × distance			✓	
Job title × Firm × 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. $\mathbb{1}$(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. 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 A.9: Robustness: Effect of Distance on Application by candidates alternate period (2019)

	(1)	(2)	(3)
1-50	–0.009***	–0.010***	–0.010***
	(0.001)	(0.001)	(0.001)
50-100	–0.021***	–0.022***	–0.022***
	(0.000)	(0.000)	(0.000)
100-500	–0.024***	–0.025***	–0.024***
	(0.000)	(0.000)	(0.000)
$>$ 500	–0.028***	–0.030***	–0.029***
	(0.000)	(0.000)	(0.000)
1-50 x $\mathbb{1}$(Non-metro)	–0.018***	–0.019***	–0.019***
	(0.002)	(0.003)	(0.002)
50-100 x $\mathbb{1}$(Non-metro)	–0.040***	–0.041***	–0.040***
	(0.001)	(0.001)	(0.001)
100-500 x $\mathbb{1}$(Non-metro)	–0.041***	–0.041***	–0.041***
	(0.001)	(0.001)	(0.001)
500 x $\mathbb{1}$(Non-metro)	–0.040***	–0.040***	–0.040***
	(0.001)	(0.001)	(0.001)
N	161293860	162497861	162505938
Mean Y	0.006	0.006	0.006
Mean Y (Same city)	0.036	0.036	0.036
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. $\mathbb{1}$(Non-metro) is a binary variable that takes the value 1 if it's non-metro job location, i.e. all cities excluding Delhi, Mumbai, Bengaluru, Chennai, and Kolkata. Job controls include education and experience required. Candidate controls include gender, education, age and age square. 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. Note that it is on account of too many observations we choose three months for alternate period.

Table A.10: Robustness: Effect of Distance on Application - Heterogeneity by population

	(1)	(2)	(3)
1-50	–0.012***	–0.024***	–0.013***
	(0.001)	(0.002)	(0.002)
50-100	–0.039***	–0.053***	–0.044***
	(0.000)	(0.000)	(0.000)
100-500	–0.045***	–0.060***	–0.051***
	(0.000)	(0.000)	(0.000)
$>$ 500	–0.049***	–0.064***	–0.054***
	(0.000)	(0.000)	(0.000)
1-50 $\times$ $\mathbb{1}$ (Below 95th)	–0.377***	0.018***	–0.383***
	(0.006)	(0.004)	(0.006)
50-100 $\times$ $\mathbb{1}$ (Below 95th)	–0.390***	0.004***	–0.394***
	(0.005)	(0.001)	(0.005)
100-500 $\times$ $\mathbb{1}$ (Below 95th)	–0.400***	–0.005***	–0.404***
	(0.005)	(0.000)	(0.004)
$>$500 $\times$ $\mathbb{1}$ (Below 95th)	–0.400***	–0.005***	–0.404***
	(0.005)	(0.000)	(0.004)
N	164251691	165964571	166030192
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. $\mathbb{1}$ (Below 95th) takes the value 1 if the population of the job city is less than 95th percentile. Job controls include education and experience required. Candidate controls include gender, education, age and age square. 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 A.11: Robustness: Effect of Distance on Application by Candidates - Additional controls

Candidate Proportion	Metro (1)	Non-metro (2)	Metro (5)	Non-metro (6)	Metro (3)	Non-metro (4)
1-50			–0.022***	–0.175***		
			(0.002)	(0.008)		
50-100			–0.055***	–0.201***		
			(0.001)	(0.002)		
100-500			–0.057***	–0.212***		
			(0.001)	(0.002)		
> 500			–0.060***	–0.216***		
			(0.001)	(0.002)		
Proportion	–0.121***	–0.504***	–0.066***	–0.783***	–0.108***	–0.413***
	(0.004)	(0.016)	(0.003)	(0.019)	(0.004)	(0.017)
1-50 x Proportion	0.055***	0.510	0.048***	1.505***	0.039***	0.208
	(0.011)	(0.390)	(0.007)	(0.352)	(0.013)	(0.170)
50-100 x Proportion	0.099***	0.497***	0.100***	0.821***	0.107***	0.421***
	(0.007)	(0.019)	(0.004)	(0.022)	(0.007)	(0.021)
100-500 x Proportion	0.127***	0.508***	0.085***	0.858***	0.113***	0.412***
	(0.004)	(0.016)	(0.003)	(0.019)	(0.004)	(0.017)
>500 x Proportion	0.130***	0.511***	0.072***	0.798***	0.116***	0.412***
	(0.004)	(0.016)	(0.003)	(0.019)	(0.004)	(0.017)
N	102007865	64546181	101776963	64419711	101775159	64417222
Controls						
Job FE	✓	✓	✓	✓	✓	✓
Candidate FE	✓	✓	✓	✓	✓	✓
Month-year FE	✓	✓	✓	✓	✓	✓
City × Job role × Dist. group	✓	✓				
Firm × Job role			✓	✓		
Firm × Title × Dist. group					✓	✓

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. Proportion variable sums up to 1 across job locations for each job role for a given candidate within the search spell. 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.

Table A.12: Robustness: Effect of Distance on Application by Candidates - Heterogeneity by Job Role Concentration

Static Proportion	(1)	Metro (2)	Non-metro (3)
1-50	−0.050***	−0.024***	−0.263***
	(0.002)	(0.002)	(0.004)
50-100	−0.090***	−0.058***	−0.295***
	(0.001)	(0.001)	(0.003)
100-500	−0.101***	−0.060***	−0.312***
	(0.001)	(0.001)	(0.003)
> 500	−0.104***	−0.064***	−0.316***
	(0.001)	(0.001)	(0.003)
1-50 x Proportion	0.188***	0.069***	2.908***
	(0.007)	(0.008)	(0.167)
50-100 x Proportion	0.249***	0.135***	2.028***
	(0.006)	(0.006)	(0.028)
100-500 x Proportion	0.275***	0.114***	2.160***
	(0.004)	(0.002)	(0.026)
>500 x Proportion	0.258***	0.107***	2.125***
	(0.004)	(0.002)	(0.026)
N	166555050	102007876	64546823
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. Proportion is share of the job role such that adds up to one across job locations. 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 A.13: Robustness: Effect of Proportion on Application by Candidates (Other Characteristics)

Candidate Proportion	Social Media Connectivity		Language Proximity	
	Metro (1)	Non-metro (2)	Metro (3)	Non-metro (4)
1-50	–0.028*	1.315***	–0.019***	–0.051***
	(0.015)	(0.032)	(0.003)	(0.016)
50-100	–0.016**	1.527***	–0.052***	–0.078***
	(0.008)	(0.024)	(0.001)	(0.010)
100-500	0.012**	1.656***	–0.051***	–0.071***
	(0.005)	(0.022)	(0.001)	(0.010)
>500	0.041***	1.666***	–0.050***	–0.074***
	(0.005)	(0.021)	(0.001)	(0.010)
Characteristic	0.010***	0.161***	0.024***	0.199***
	(0.000)	(0.002)	(0.001)	(0.013)
Proportion	–0.058***	–0.605***	–0.076***	–0.645***
	(0.003)	(0.017)	(0.003)	(0.023)
1-50 x Proportion	0.026***	1.376***	0.053***	1.410***
	(0.007)	(0.355)	(0.007)	(0.350)
50-100 x Proportion	0.088***	0.661***	0.107***	0.670***
	(0.005)	(0.020)	(0.005)	(0.026)
100-500 x Proportion	0.076***	0.645***	0.095***	0.712***
	(0.003)	(0.017)	(0.003)	(0.023)
>500 x Proportion	0.062***	0.618***	0.083***	0.647***
	(0.003)	(0.017)	(0.003)	(0.024)
1-50 x Characteristic	0.002	–0.121***	–0.008**	–0.172***
	(0.002)	(0.003)	(0.004)	(0.017)
50-100 x Characteristic	–0.002**	–0.142***	–0.008***	–0.165***
	(0.001)	(0.002)	(0.002)	(0.014)
100-500 x Characteristic	–0.005***	–0.157***	–0.013***	–0.198***
	(0.000)	(0.002)	(0.001)	(0.013)
>500 x Characteristic	–0.009***	–0.158***	–0.011***	–0.206***
	(0.000)	(0.002)	(0.001)	(0.013)
N	101755699	64204499	102007876	64546823
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. Social Media Connectivity is measure of social connectedness between two locations, therefore, measures the relative probability of a Facebook friendship link between a given Facebook user in two locations. Here we use log(SCI). Language proximity gives the measure of language overlap between the two states, given by Language Proximity$_{o,d} = \sum_{m=1}^{23} \left(s_{o,m} \times s_{d,m}\right)$ where $s_{o,m}$ is the share of speakers of language m in state o, $s_{d,m}$ is the share of speakers of language m in state d.Proportion variable sums up to 1 across job locations for each job role for a given candidate within the search spell. 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 2020âDecember 2020.

Table A.14: Robustness: 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)
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 A.15: Robustness: Effect of Distance on Shortlisting by Employer: Heterogeneity

	(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)
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.

Table A.16: Applications per Vacancy: 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)
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.

B Appendix: Other Details

B.1 Additional 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.

Appendix 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 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. Appendix 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

[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.

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 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. Appendix 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.

B.2 Distaste: Role of Posted Wages and Other Characteristics

We consider the role played by various factors in influencing the candidate's distaste for distance. 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 yields around 62 million candidate×job observations with an average application rate of 1.7%.[2]

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

[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 available on request.

(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 the candidate (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.

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 × ln (wage)		0.028***			
		(0.003)			
50-100 × ln (wage)		0.023***			
		(0.001)			
100-500 × ln (wage)		0.022***			
		(0.001)			
> 500 × ln (wage)		0.022***			
		(0.001)			
1-50 × $\mathbb{1}$(wage > med)				0.022***	0.023***
				(0.004)	(0.004)
50-100 × $\mathbb{1}$(wage > med)				0.020***	0.019***
				(0.001)	(0.001)
100-500 × $\mathbb{1}$(wage > med)				0.021***	0.021***
				(0.001)	(0.001)
> 500 × $\mathbb{1}$(wage > med)				0.020***	0.020***
				(0.001)	(0.001)
N	62515398	62575521	62515398	62575521	51419446
Mean Y	0.018	0.018	0.018	0.018	0.018
Mean Y (Same city)	0.108	0.108	0.116	0.116	0.141
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. Ln(wage) is log wage offered by the job. $\mathbb{1}$(wage> med) is a binary variable takes the value 1 if the offered wage is above median. 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: Robustness: Effect of Distance on Application by candidates - Distaste by Gender, Education and Experience level

	(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)
N	166555050	166415759	135272680
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. Female, Graduate, and Experienced are binary variables that take value 1 if the candidate is female, education level is graduate and above, and experience in years is more than 1 year, takes value 0 otherwise. 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.

B.3 Institutional Details: UDAN Scheme

Annually India's air passenger count is more than 300 million, still large part of this is concentrated in handful of bigger cities. In 2017, UDAN scheme was launched by Government of India to increase connectivity between the smaller cities. Through subsidy offerings for airlines and reduced airfare costs the scheme seeks to promote regional connectivity, and foster economic development in these areas. As part of the scheme, 451 routes were operationalized as of January 2022. UDAN has played an important role in expanding air travel access beyond major metropolitan cities, to help bridge the connectivity gap in India.

Using introduction of UDAN routes as natural experiment, we analyze the impact of the scheme on job applications on connected routes. We consider the total number of applications from candidates between June 2018-December 2019. Among the connected UDAN cities under the scheme, we have selected 54 candidate cities as treatment cities basis the airports were smaller – excluded older airports like Patna[4], removed the chopper routes, and sea routes. Next, using the number of applications between candidate and job city on a given date, we create a potential application set assuming a job could potentially receive applications from each of these 54 treatment cities and then measure the impact of the introduction of UDAN routes using Sun & Abraham (2021) estimator. There were 55 UDAN routes in the period of analysis, compared to 100 routes that were introduced post June 2018[5] for the included 54 treatment cities. There were 1692 candidate-job city locations[6] that were not connected as part of UDAN scheme in our sample. Please see Figure B.1 for detailed breakup of routes by distance categories and by quarter. Many more UDAN routes were introduced in the 2019, compared to 2018 in our analysis period. Conversely, as one would expect we see larger number of connections being introduced for the medium length routes that is more than 100 miles.

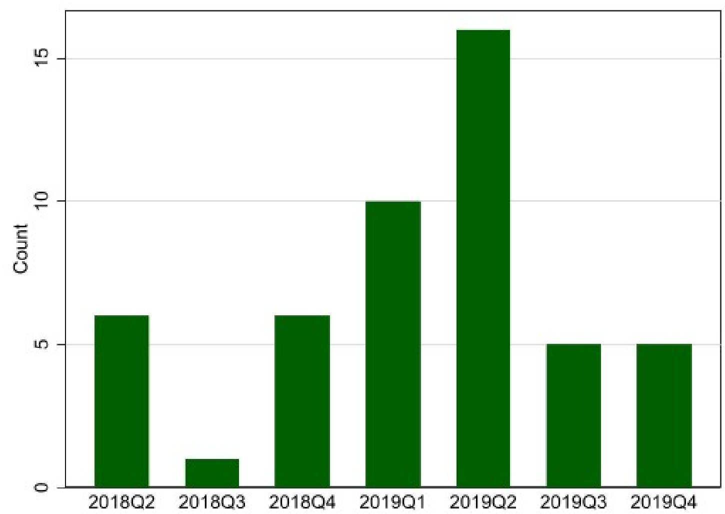
[4]The cities that already had an airport and excluded from candidate city's list were – Ahmedabad, Amritsar, Bangalore, Bhubaneswar, Chandigarh, Chennai, Delhi/NCR, Dibrugarh, Hyderabad, Imphal, Jaipur, Kochi, Kolkata, Lucknow, Mumbai, Patna, Pune, Thiruvananthapuram, and Varanasi.

[5]Routes introduced till June 2022

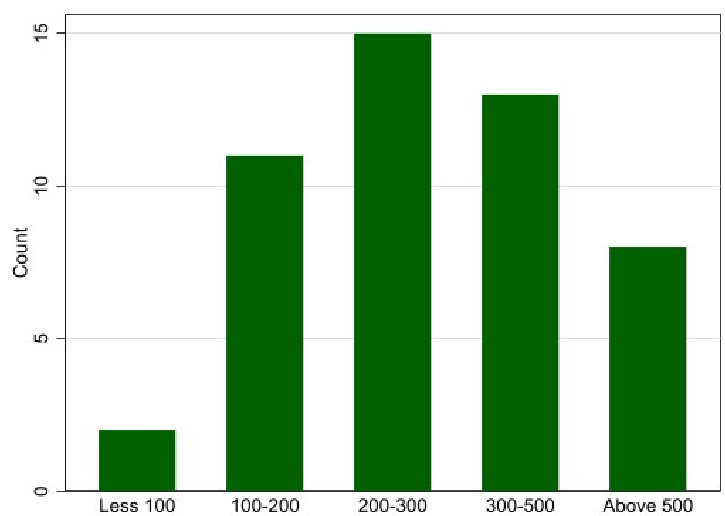
[6]The ij and ji connections counted separately, origin and destination matters.

Figure B.1: *UDAN Sample*: Air route details

(a) Treated routes by introduction quarter



(b) Treated routes by distance categories



Notes: We plot the air routes that are treated in our analysis period June2018-December2019. Distance groups (in miles) are defined on the basis of distance between the posted job location city and candidate location city. We have excluded the routes that were introduced before or after the period of analysis. Also, exclude the chopper routes introduced during this period as part of UDAN scheme.