I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17105

ICT Skills and Labor Market Outcomes in India: Evidence from Cell Tower Expansion

Mehtabul Azam
M. Shahe Emran
Forhad Shilpi

JUNE 2024

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17105

ICT Skills and Labor Market Outcomes in India: Evidence from Cell Tower Expansion

Mehtabul Azam
Oklahoma State University and IZA

M. Shahe Emran
IPD, Columbia University

Forhad Shilpi
DECRG, The World Bank

JUNE 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

ICT Skills and Labor Market Outcomes in India: Evidence from Cell Tower Expansion*

Using a nationally representative large-scale survey of individual ICT skills in India (Multiple Indicators Survey, 2020), we provide evidence on the effects of ICT skills on labor market outcomes and household welfare as measured by per capita expenditure. We study the effects both at the extensive and intensive margins of labor market. To tackle the challenges in identification arising from unobserved individual heterogeneity in the acquisition of ICT skills, we develop an instrumental variables (IV) strategy. The IV approach exploits the dramatic expansion of cell towers in India as a source of supply-side variation, and relies on an institutional feature of the telecom market, the "telecom circles", to construct a leave-own-out instrumental variable. The evidence suggests no significant effects at the extensive margin (null effect on both labor force participation and employment). In contrast, there are important effects at the intensive margin: a 10 percentile higher ICT skills index increases the probability of salaried employment by 6.5 percentage points, and leads to a 9.5 percent higher per capita expenditure. Employment transitions happen from daily wage employment and self-employment to salaried employment. The effects vary substantially across gender: women face a penalty in the form of a lower impact on salaried employment, but the impact on per capita expenditure is larger for households with ICT-skilled women. The higher impact on women reflects the fact that the family stock of ICT skills is much larger in households with ICT-skilled women. Contrary to conventional wisdom, the lower caste individuals enjoy a larger positive effect on salaried employment, despite their unfavorable labor market network, possibly due to the employment quotas in public sector employment. The impacts on salaried employment and per capita expenditure for Muslims are comparable to the other social groups. We find no significant rural-urban differences.

JEL Classification: J24, J62, I30, O12, O15

Keywords: ICT skills, labor market, employment, salaried employment, daily wage employment, self-employment, per capita expenditure, gender, caste, religion, towercos, telecom circles, India

Corresponding author:
Mehtabul Azam
Department of Economics
Oklahoma State University
Stillwater
Oklahoma 74078
USA

E-mail: mazam@okstate.edu

* We thank Hanchen Jiang for comments on an earlier draft. The standard discalimers apply.

(1) Introduction

Information and communications technology skills (ICT skills) are considered by many as necessary for better labor market opportunities and outcomes, both in developed and developing countries. The 2016 World Development Report of the World Bank "Digital Dividends" calls ICT skills "foundational", and International Labor Organization (ILO) considers ICT skills (basic digital skills) part of core skills for the labor market in the 21st century (Aggarwal (2021)). Many governments in developing countries have identified digital infrastructure expansion as an important part of their development strategy, with India and China at the forefront. In India, access to cell phone service has increased exponentially in the recent decades; teledensity increased from 3.58% in 2001 to 88.66% in 2020 with 1176.79 million telephone subscribers. Although access to internet lagged far behind for most of this period, it witnessed explosive growth in the more recent years: 63% of the telephone subscribers also had internet connection in 2020.[2] The development of ICT skills in India, however, remains low: in 2020-21 almost 75 percent of Indian population did not have very basic ICT skills such as how to copy or move a file or folder.[3] If having ICT skills is important for employment in formal sector including the public sector, then the recent expansion of access to internet in India has failed to improve the economic prospects for most people, and some argue that the digital revolution in India is widening economic inequality (Bajpai and Biberman (2019)).

The argument that ICT skills confer an edge in the labor market seems plausible given the spread of computers in business, industry, public sector, and a spectacular growth of online trading. This plausible argument, however, lacks an empirical foundation in the context of developing countries, including India. We are not aware of any studies that provide evidence on the effects of ICT skills on labor market outcomes in a developing country. While there is a substantial experimental literature on the effects of job training and job search interventions which often include teaching ICT skills along with other vocational skills, we are not aware of any studies that isolate the ICT skills component of the intervention package.[4] Our study differs from the experimental literature in two other

[2]Source: Government of India (2020) Telecom Statistics India-2020, Department of Telecommunications, Government of India.

[3]In rural areas 82 percent population did not have such ICT skills. Source: Government of India (2023).

[4]For recent surveys of the experimental literature, see Carranza and McKenzie (2024), and Caria et al.

important respects. First, most of the experimental studies focus on the younger cohorts, especially recent graduates from vocational training or college, thus do not capture a broad cross-section of the society as we do. We also provide a comparative analysis of the effects of ICT skills across young (21-40 years old) and (41-60 years) old cohorts. Second, many experimental analysis study the transition from unemployment to employment, and do not analyze the intensive margin, i.e., the occupational transitions among the employed. We study the effects of ICT skills both at the extensive margin and the intensive margin. Third, our analysis takes advantage of a large data set that covers all districts in India, and thus the estimates refer to representative national population. There is a small but growing literature that studies the employment effects of the arrival or expansion of high-speed internet in Africa, but they do not estimate the effects of ICT skills of individuals on their labor market outcomes (see, for example, Chiplunkar and Goldberg (2022)).

This paper provide the first evidence on these issues in the context of India, a country of 1.4 billion people in 2020 that achieved high internet expansion through mobile phones in the decades of the 2000s and the 2010s. Our focus is on the following interrelated questions. Does ICT skills make a person more employable? Does ICT skills open up employment opportunities in higher-paid and stable jobs such as salaried employment? Does it help people to graduate from the informal daily wage employment that still account for a significant share of employment in India? What are the distributional consequences of digital transformation in Indian economy? Are the relatively disadvantaged segments of the society such as lower castes left behind while people from advantageous socioeconomic background (higher caste, men, college educated) reap most of the benefits?

Note that ICT skills can affect labor market outcomes in a variety of ways: (i) providing necessary skills for the ICT-intensive modern sectors, (ii) through online job search, and (iii) a positive impact on educational outcomes through internet-based educational resources. Our goal is to estimate the total effects of ICT skills capturing all these different aspects. We also provide evidence on the importance of the education and job search channels.

Our empirical analysis is based on data from the Multiple Indicator Survey (MIS) collected by the Indian National Sample Survey Organisation (NSS) in 2020-2021. The survey is the first data source to provide information on ICT skills for a large sample of

(2024).

nationally representative Indian population. Our estimation sample consists of 622,764 individuals in age group 21-60 covering all the 640 districts of the Indian 2011 Census. To develop an identification strategy, we rely on an instrumental variables design that exploits two institutional features of the telecom market in India: (i) the dramatic expansion of cell towers and cell phone subscription, facilitated by a complementary government regulatory policy, (ii) a particular feature of telecom regulation policy: 22 telecom circles introduced by the National Telecom Policy (NTP) of 1994.[5]

The main challenge for identifying the effects of ICT skills on labor market outcomes comes from unobserved heterogeneity in the demand side which affects both the acquisition of ICT skills and the labor market outcomes of an individual. We thus rely on exogenous variation in the supply side of ICT: the dramatic growth in cell towers in India in the last two decades.[6] Since most of the internet access in India is on mobile phones, cell tower expansion is an important determinant of internet access and ICT skills. The number of cell towers (Base Transceiver Station, BTS for short) grew dramatically in India from 178,000 in 2008 to 775,000 in 2015, and to 22,116,000 in 2020.[7] Such a dramatic expansion implies that the endogeneity that arises because of targeting the towers to districts with high economic endowment is unlikely to be a major issue in the context of India.

This extraordinary expansion of cell towers was the result of a unique combination of factors. First, it was spearheaded by Towercos, private companies set up to build and maintain cell towers, and to provide the "passive infrastructure" which is rented and shared by different telecom operators for their "active" broadcast equipment such as antennas.

[5]The term a "cell tower" in this paper refers to a Base Transceiver Station (BTS), and the number of cell towers equals the number of BTS. This reflects two considerations. First, the availability and quality of cell phone service in a location depends on the number of BTS, and not the number of towers (physical structure) because one tower can host multiple BTSs for different spectrum, and, even for a given spectrum, different BTSs for different telecom operators. Second, our data source for cell towers in India is open source repository OpenCellID, which defines a cell tower as a BTS. For more details on the cell tower data we use, please see section (4.3) below.

[6]The insight that supply side variations can help deal with demand side unobserved heterogeneity and endogeneity is widely-used in the literature. See the discussion by Angrist and Krueger (2001). Some studies use spatial and temporal variation in lightning frequency (flash rate) for identifying the effects of the arrival of fast internet. Andersen et al. (2012) was the first to establish an empirical link between flash rate and IT diffusion in the context of USA, and Chiplunkar and Goldberg (2022) used this for Africa to analyze the effects of 3-G internet on structural change in employment at the district level. However, we cannot adopt this approach because lightning variation has little power in explaining the cell tower expansion in the context of India. We provide a fuller discussion with the relevant evidence in the online appendix.

[7]The number of macro towers also experienced high growth: from 250,000 in 2007 to 606,300 in 2020, with 29,000 new macro towers built every year on an average (Gupta et al. (2020)).

Government Of India (GOI) created a regulatory framework that imposes very little costs of entry into Towerco industry; there are no licensing or access regulation in India, only business regulation.[8] As a result, 84 percent of towers in India are managed by Towercos (Houngbonon et al. (2021)).[9] Second, the "passive infrastructure sharing" model made entry by the competing telecom operators in a location much easier, and different operators aggressively expanded their penetration across different telecom circles.[10] Third, from 2016 to 2018, Reliance Jio Phone of the conglomerate Reliance Industries Limited invested $35 billion to expand its cell phone services, constructed 135,000 new 4G cell towers, and effectively "blanketed" the whole country with cheap cell phone service, offered at a quarter of the industry average price before 2016 (Purnell (2018)).

However, one might worry that, even with such dramatic expansion of towers and cell service, cell tower density in a district might be correlated with it's economic potential and structure, because Towercos would build more towers per capita in districts with more growth prospects. If this is the case, the exclusion restriction imposed on the tower density in a district in an IV approach is violated, because tower density in a district acts as a proxy for its economic structure and growth prospects. To tackle this, we develop a leave-own-out (LOU) instrumental variable constructed as follows. We regress the tower density in a district on the tower density in *all other districts in a telecom circle* as defined by the National Telecom Policy (NTP) of 1994.[11] The predicted tower density from this regression is used as an IV (henceforth called "LOU IV") in a just-identified model, after conditioning on state-region fixed effects and a rich set of district and individual level controls.[12] The telecom circles provide a natural spatial unit to define the LOU IV because the quality (upload and download speeds, for example) and costs across districts within a circle are likely to be correlated as they are served by the same operators under the same pricing

[8] In contrast, countries such as Pakistan and Egypt require licensing, access regulation and business regulation.

[9] In comparison, in USA and Canada, about 70 percent of the towers are built and managed by Towercos. Only China has more Towercos than India, almost 100 percent towers being managed by the Towrcos.

[10] In countries such as Pakistan and Egypt where Towerco industry is not developed, the telecom operators themselves need to incur huge capex (capital expenditure) to build the passive infrastructure. Since the towers built by one operator is usually not rented to the other competing operator, this increases the capital cost of providing cell services, limiting their profitability only in urban and more prosperous rural districts.

[11] We allow for heterogeneity across telecom circles, and with respect to economic characteristics of a district when constructing the instrumental variable.

[12] For discussions on the advantages of a just-identified IV model, see Angrist and Kolesár (2023) and Kolesár et al. (2015).

policies.[13]

It is important to appreciate that our identification strategy is different from the standard leave-one-out spatial instruments used, for example, by Acemoglu et al. (2008), Persson and Tabellini (2009), Fruehwirth et al. (2019) and Acemoglu et al. (2019). The standard approach uses a *weighted or simple average of the endogeneous variable in a suitable spatial unit* (leaving the relevant observation out) as an instrument. For example, Fruehwirth et al. (2019) use the religiosity of the peer group in a school as an IV for religiosity of an individual student, and Persson and Tabellini (2009) use democracy in the neighboring countries as an instrument for democracy in a country. If we follow the standard approach, we can use the average ICT skills of others in a village (excluding the concerned individual) as an instrument for ICT skills of an individual. However, the LOU average ICT skills in a village is likely to be endogenous because of peer effects in acquisition of ICT skills, giving rise to the "reflection problem" of Manski (1993).[14] Instead of an IV based on ICT skills of the neighbors or a peer group, our approach relies on the supply side variation due to the dramatic expansion of cell towers and cell phones in the last two decades.

A final issue for our empirical strategy is that the availability of towers may spur growth of ICT-intensive economic activities in a district through firm entry and expansion, thus increasing the demand for ICT skills by firms in that district. In this case, the estimated effect using the LOU IV based on cell towers captures some of the firm-side effects along with the effects of acquiring ICT skills at the individual level.[15] To estimate the effects of acquiring ICT skills at the individual level, we control for the changes in employment structure in a district over the relevant decades.

Note that our LOU IV varies across districts, and we cannot use district fixed effects to mop up any residual correlation between the LOU tower density and unobserved district characteristics. We adopt a two-pronged strategy to deal with this issue: (i) include state-region fixed effects, and (ii) include a set of district characteristics from 2011 census, which

[13] Indian government allocated the telecom spectrum auctions based on these telecom circles. Note that the same operator charges different rates in different circles, especially the charge is likely to be lower in districts where the internet speed offered is lower.

[14] It is easy to see the refection problem in the two person case. Whether individual i invests in acquiring ICT skills depends on what individual j does, and vice versa. For an excellent discussion on the limitations of the standard LOU spatial IV approach, please see Betz et al. (2018).

[15] Only part of the firm-side effects which is correlated with the ICT skills in a district is captured.

are predetermined for our analysis.[16] The state-region fixed effects deal with potential spatial correlation, for example, because of common agroecological factors shared by districts in a region. We provide evidence that the state-region fixed effects and district level controls together act as a sufficient statistic for district fixed effects. As an additional check, we implement the IV bounds approach of Conley et al. (2012) to ensure that our conclusions are robust even if LOU IV exerts a small direct impact on the outcome variables, thus the exact exclusion restriction imposed in the IV estimation is locally violated.

The evidence from our empirical analysis suggests three key conclusions. First, at the national level, ICT skills substantially improve the probability of salaried employment and reduce the probability of daily wage employment and self-employment. In contrast, there is no perceptible impact at the extensive margins: labor force participation and employment. Having ICT skills also increases per capita household consumption, suggesting a higher welfare. Our IV estimates suggest that a 10 percentile higher ICT skills index increases the probability of regular salaried employment by 6.5 percentage points, and reduces the probability of self employment by 3.6 percentage points and that of daily wage employment by 3.4 percentage points. The evidence suggests a 9.5 percent higher per capita expenditure when the ICT skills index is 10 percentile higher. Second, when we allow for local violation of the exclusion restriction imposed on the LOU tower density as an instrument, the estimated impacts remain robust: none of the reported 90% confidence interval bounds from the Conley et al. (2012) approach include zero for all the outcomes of interest. Third, there are important heterogeneity across gender, caste, religion, and age cohorts (young vs. old), but no significant differences are found in the causal effects across rural and urban samples. The evidence suggests a gender penalty against women in the labor market: while ICT skills help a woman to move away from daily wage employment and self employment to salaried employment, the magnitude of the effect on salaried employment is about 50% lower compared that for men. In contrast, households with women who have ICT skills enjoy higher per capita consumption. We find suggestive evidence that the higher impact for women on per capita expenditure reflects assortative matching in marriage: the stock of family ICT skills (husband+wife) is substantially higher in those households where women have high ICT skills. Somewhat unexpectedly, we find that the labor market effects are

[16] As a robustness check, we use 2001 district characteristics.

much stronger for the Scheduled Castes (SC) and Scheduled Tribes (ST): a 10 percentile higher ICT skills index implies a 2.8 percentage points higher probability of getting a salaried employment for Scheduled Tribes, and a 3.5 percentage points higher probability for Scheduled Castes. Since SC and ST lack labor market network for salaried employment, this larger impact compared to all other social groups including the upper castes in India may seem surprising. However, employment quotas for SC and ST in public sector employment may explain this evidence as there are a lot more salaried employment vacancies available for the SC and ST. Evidence indicates that Muslims, the most disadvantaged social group in India in many respects, enjoy comparable gains in salaried employment and per capita expenditure once they acquire ICT skills. But the negative impact of ICT skills on daily wage employment is much stronger for Muslims, while the negative impact on self-employment is smaller. This suggests that ICT skills lead to a different employment transition for Muslims: they not only get more salaried employment, but also move away from daily wage employment to self-employment.

We provide evidence on the role played by education and job search channels in mediating the effects of ICT skills on labor market outcomes. We find that education is not important as a mechanism, while there is suggestive evidence that job search channel plays a small positive role. The evidence taken together suggests that ICT skills affect labor market outcomes primarily by enhancing the employability of an individual.

The rest of the paper is organized as follows. In section 2, we provide a discussion on the related literature. Section (3) discusses the institutional context with a focus on the role of towercos and government policies in the dramatic expansion of cell towers in India in recent decades, while section (4) is devoted to a description of the data and summary statistics. We discuss the identification challenges and develop the empirical strategy in section (5). Section (6) organized in a number of subsections report the empirical estimates, with the OLS estimates in subsection (6.1), IV estimates in subsection (6.2), and the estimated Conley et al. (2012) bounds in subsection (6.3). The evidence on heterogeneity in the effects of ICT skills across gender, caste, religion, age cohorts (young vs. old), and location are reported in section (7). Section (8) explores the mechanisms: the role of education and job search as mediating channels. The paper concludes with a summary of the main findings in section (9).

(2) Related Literature

As noted in the introduction, we are not aware of any studies that estimate the effects of an individual's ICT skills on his/her labor market outcomes and household expenditure in the context of developing countries. Moreover, there is only a limited literature on the labor-market effects of ICT skills at the individual level in the context of developed countries. To our knowledge, only Falck et al. (2021) and Non et al. (2021) look at the individual's ICT skills and labor market outcomes.[17] Falck et al. (2021) use the 2012 Programme for the International Assessment of Adult Competencies (PIAAC) data from 19 OECD countries and exploit technologically induced variation in Internet availability across countries and across small geographical areas within a single country, Germany. They find that a one standard deviation increase in digital skills increases wages by 24 (internationally) to 31 (for Germany) percent. Their IV estimates are about twice as large as corresponding OLS estimates. Non et al. (2021) combine the 2012 PIAAC score for Netherlands with register data from Netherlands on labor market outcomes in 2012-2019. Using OLS, they find that a one standard deviation increase in digital skills is associated with an increase in wages by four to six percent.

(3) Institutional Context: Towercos, Base Transceiver Stations (BTS), and Expansion of Cell Phones and Internet in India

In 2001, the total number of telephone subscribers in India was 36.28 million with a teledensity of 3.58%. Out of those 36.28 million telephone subscribers, only 3.58 million were mobile subscribers (32.70 million landline subscribers). In 2020, the total number of telephone subscribers in India stood at 1176.79 million with a teledensity of 88.66%. Only 19.13 million of telephone subscribers had the landline in 2020. Hence, the role of the landline is virtually negligible in the telecom subscriber base. Similarly, in 2020, there were a total of 743.19 million internet subscribers in India, out of which only 22.42 million

[17]There is a small but growing literature that estimates the effects of broadband or internet access on labor market, but they do not analyze the role of ICT skills of individuals (for an excellent survey, see Hjort and Tian (2024)). Chiplunkar and Goldberg (2022) provide evidence on the effects of 3-G internet expansion on employment and sectoral employment composition at the district/commune level in Africa. Hjort and Poulsen (2019) use the gradual arrival of internet cables on the coast to study the effects of fas internet on employment and firm entry and productivity in 12 African countries. Zuo (2021) use quasiexperimental variation in Comcast broadband in USA to study the effects on labor market outcomes of low income families, and finds a positive impact on labor force participation and a lower probability of unemployment with access to discount broadband access.

had wired internet connections. Given the dependency of internet/phone access on wireless technology, the availability of cell tower and cell tower density in a location play a key role in access to internet.[18]

When building a telecommunications network, different types of structures are used to set-up the Base Transceiver Stations (BTS) and antennas which include the so called "macro towers", i.e., the tall towers erected for the sole purpose of telecommunications infrastructure, and the "small cell sites" such as utility poles and roof tops used for setting up BTS and antennas. A cell tower, especially a macro tower, usually houses multiple BTS and many antennas, each BTS dedicated for a specific spectrum. Since access to cell service depends on the number of BTS rather than the number of macro towers, our analysis focuses on the number of BTS as a measure of supply of cell service. Following the recent literature, we count the number of "cell towers" by the number of BTS. Thus, we use the term "cell tower" and BTS interchangeably.

Traditionally, the cell towers were built and managed by the telecom operators which meant that they had to incur huge amounts of capital expenditure. This also had important implications for their entry decisions: they targeted primarily the most profitable segments of a market to ensure high enough returns to recover the fixed costs. Such entry decisions make the spatial distributions of cell towers across districts endogenous. The Towerco model freed the telecom operators from the huge financial burden of building their own towers, especially macro towers which require substantial amount of capital expenditure. Towercos are private companies set up to build and maintain cell towers, and provide the "passive infrastructure" which is rented and shared by different telecom operators for their "active" broadcast equipment such as antennas. Once a tower is built, it is shared by multiple cell operators and other connectivity operators such as microwave networks, TNT, and radio, thus spreading the burden of fixed costs. The number of cell towers (BTS) grew dramatically in India from 178,000 in 2008 to 775,000 in 2015, and to 22,116,000 in 2020. The number of macro towers also experienced high growth: from 250,000 in 2007 to 606,300 in 2020, with 29,000 new macro towers built every year on an average (Gupta et al. (2020)). This extraordinary growth in cell towers was facilitated by a regulatory framework highly conducive to entry into Towerco industry: there were no licensing or access regulation in

[18] For discussions on the evolution of teledensity and cell phone subscriptions in India, see Mangla and Singh (2021) and TRAI and NCAER (2012).

India, only business regulation. In contrast, countries such as Pakistan and Egypt imposed licensing, access and business regulation on a Towerco, and stifled their growth. In 2020, 84 percent of towers in India were managed by Towercos (Houngbonon et al. (2021)).

The "passive infrastructure sharing" model of Towercos has been hugely successful in providing macro towers in locations which would not be profitable for a single cell phone operator. Without the financial burden of building their own towers, many telecom operators aggressively expanded their penetration across different states and districts within a state. The entry of Reliance Jio in the cell phone market provides a striking example. In 2014, Reliance Jio Infocomm Ltd of the Reliance Industries, made an agreement with Indus Tower, the largest Towerco in India, to share Indus's 113,490 towers to offer Jio Phone services. A latecomer to the cell phone market in India, Jio Phone entered into a large part of the national market comprising of 15 of India's 22 telecom circles served by Indus towers in 2016, without investing anything in building towers. Reliance Jio then aggressively expanded their cell service by building their own 4G towers, and by September 2020, they had 135,000 macro towers, virtually blanketing the whole country. Reliance Jio offered cheap phones and huge discounts (on average a 75% discount on the competitors' price) to expand their subscription base; in 4 years they had 393 million cell phone subscribers (34 percent of the market).[19] The important point for our empirical analysis is that such dramatic expansion of cell towers and cell service at low costs implies that almost all districts, irrespective of their economic endowments and prospects, had access to cell services, and not only the rich, but also middle class and relatively poor families could afford cell service and data access.[20]

(4) Data and Variables

(4.1) Data

We use the Multiple Indicator Survey (MIS) collected by the Indian National Sample Survey Organistaion (NSS). The survey was initially planned to be conducted during January–December 2020, but due to the Covid-19 pandemic, the data collection was continued up to 15.08.2021 (Government of Inida, 2023). The survey collected information from 276,409 households (164,529 in rural areas and 111,880 in urban areas) for about 1163,416 individuals

[19]In a Wall Street Journal Article on Reliance Jio in 2018, Purnell wrote "Mr. Ambani wanted to build a network that would also cover more than 18,000 cities and towns and 200,000 villages, touching some places that didn't have electricity yet."

[20]Reliance Jio offered free calls and data for the first 3-6 months.

(713,501 in rural areas and 449,915 in urban areas). The sample is based on stratified multi-stage sampling design, and all the empirical analysis in the paper uses the survey weights provided in the data. We restrict our sample to individuals in the age group 21-60 years who are not in school. The age group 20-60 approximates the working age in India as most of the public sector jobs have retirement age of 60. This leaves us with a sample of 622,764 individuals covering all the 640 districts of the Indian Census 2011.[21] Summary statistics for our main variables are reported in Table 1.

(4.2) Information, Communication, and Technology (ICT) skills

The survey inquires for persons of age 15 years and above a) whether able to copy or move a file or folder; b) whether able to use the copy and paste tools to duplicate or move information within a document; c) whether able to send e-mails with attached files (e.g. document, pictures and video); d) whether able to use basic arithmetic formulae in a spreadsheet; e) whether able to connect and install new devices (e.g. modem, camera, printer; f) whether able to find, download, install and configure software; g) whether able to create electronic presentations with presentation software (including text, images, sound, video, or charts); h) whether able to transfer files between a computer and other devices; i) whether able to write a computer program using a specialized programming language. These are self-reported information with yes/no answers. Using these binary indicator variables from (a) to (i), we construct an ICT skills index based on a principal component analysis (PCA). The first principal component is used as the index in our analysis. We normalize the ICT skills index by subtracting the mean and dividing it by standard deviation. Since one standard deviation from the mean in ICT skills is a substantial jump in skills (about 34.1 percentile jump in ICT skills), we further multiply the normalized ICT skills index by 3.98 so that a one unit change in our ICT skills index means approximately a 10 percentile change in skills (a 0.254 standard deviation change). For assessing the magnitude of the estimated effect, we refer to a 10 percentile increase in ICT skills and all the estimates reported in this paper refer to the effects of a 10 percentile change in the ICT skills index. Table 2 provides the descriptive statistics of different components of the ICT skills.

[21] In the survey data, there are 684 districts because of the creation of new districts between 2011 and 2020. We map the new districts to the 2011 census districts.

(4.3) Cellular Towers Data

We use number of towers data from OpenCellID (https://www.opencellid.org/). OpenCellID is a huge, open-source, partially crowd-created repository of cellular towers.[22] It provides the latitude and longitude of mobile tower locations across the world. The data counts cell towers by the number of Base Transceiver Station (BTS). Thus, the cell tower infrastructure includes many types of structures such as a roof top housing of a BTS, not only the so-called "macro towers", the tall towers erected for the sole purpose of telecommunications. We downloaded the location data for India from the repository on December 23, 2023, and mapped the towers' locations to Indian districts using the district wise shape file for India based on 2011 Census districts. In the database, 2.63 million towers are found for India in December 2023. According to official data, the number of mobile base transceiver stations (BTS) were 2.40 million as of December 9, 2022.[23] The database also reports the date when the tower first appeared, and we discard towers that appeared after 2019. There are 2.36 million BTS in our sample, and we divide the number of BTS in the district by the 2011 Census population of the district to get a district tower density. In 2019, the average number of cell towers was 16.30 per 10,000 persons with standard deviation of 33.75.

(5) Empirical Issues and Identification Strategy

To understand the challenges in estimating the effects of ICT skills on labor market outcomes and household welfare with survey data, it is useful to consider the following triangular model:

$$\begin{aligned} Y_{idrc} &= \alpha_0 + \gamma ICT_{idrc} + \delta X_{idrc} + \xi_{drc} + \pi(\Delta E_{drc}) + \theta Z_{idrc} + \varepsilon_{idrc} \\ &= \alpha_0 + \gamma ICT_{idrc} + \delta X_{idrc} + \xi_{drc} + \pi(\Delta E_{drc}) + \phi_{idrc}; \qquad \phi_{idrc} = \theta Z_{idrc} + \varepsilon_{idrc} \end{aligned} \tag{1}$$

$$ICT_{idrc} = \beta_0 + \eta X_{idrc} + \xi_{drc} + \mu(\Delta E_{drc}) + \lambda Z_{idrc} + \upsilon_{idrc} \tag{2}$$

where Y_{idrc} is the outcome of interest for individual i residing in district d in state-region r in telecom circle c.[24] ICT_{idrc} is the information, communication, and technology (ICT)

[22] https://datacatalog.worldbank.org/search/dataset/0038043/Global-OpenCellID-cell-tower-map

[23] Source: GOI (2022) Press Release.

[24] The state-region is constructed by the National Sample Survey (NSS) by grouping similar contiguous districts (within the same state). In the NSS data, the 684 districts in India are aggregated into 83 state-regions.

skills index, X_{idrc} is a vector of observable individual characteristics, Z_{idrc} is a vector of unobserved (to the researcher) individual characteristics, ξ_{rdc} is district-fixed effects, $\triangle E_{drc}$ is a measure of changes in employment structure in district d over the relevant decades, and ε_{idrc} is individual-specific idiosyncratic errors.

Our main interest is parameter γ which captures the impact of ICT skills on the outcome of interest. We consider various labor market outcomes such as unemployment, and the type of employment (e.g, salaried vs. daily wage employment). We also analyze the effects of ICT skills on household welfare as measured by per capita expenditure. For vectors $X$and Z, we focus on those variables which affect both the acquisition of ICT skills (i.e., the selection equation (2)) and the outcomes of interest (i.e., equation (1)). An example of a variable in vector X in our context is education level of a person. A more educated person is more likely to acquire ICT skills as it is easier for them to learn these skills, and more education also improves the labor market outcomes. This suggests that omitting education from equation (1) is likely to bias the estimate of γ upward.[25]

The simple OLS estimate of the parameter γ from equation (1) is expected to be biased because the error term in the regression is not the idiosyncratic error ε_{idrc} but $\phi_{idrc} = \theta Z_{idrc} + \varepsilon_{idrc}$ and covariance between ICT skills index and the error term is given by $Cov\left(ICT_{idrc}, \phi_{idrc}\right) = Cov\left(\theta Z_{idrc}, \lambda Z_{idrc}\right) \neq 0$. It is in general not possible to pin down the direction of bias from ex-ante arguments. Perhaps the most widely-discussed source of omitted variable bias in equation (1) is 'ability bias' because most data sets lack data on cognitive and noncognitive abilities of an individual, and thus they are elements of the vector Z in equation (1). It is common to argue that omitted ability introduces upward bias in the OLS estimate. However, this widely-used argument is based on an implicit assumption: ability is something that increases productivity across the board, a high ability person reaps higher benefits both in ICT-intensive activities, and the alternatives (say, farming). In a model where ability is multidimensional, the ability bias can be downward. In this case, some people might be better at doing ICT-intensive work, while others are better at the alternative such as farming. People choose occupations/employments based on

[25]However, note that this argument implicitly assumes that ICT skills do not affect education level of a person in our context. We provide evidence consistent with this later in the paper. If ICT skills have significant effects on education, then this becomes a "bad control" in the terminology of Angrist and Pischke (2009).

their comparative advantage, and ability bias is downward when we use the farmers as the counterfactual for the people with ICT skills who have lower ability in farming compared to the self-selected farmers. The important point here is that, even when the only source of bias is omitted abilities, we cannot sign the direction of the bias on a priori grounds. An additional source of bias is the classical measurement errors in ICT skills index which may introduce attenuation bias.

(5.1) An Instrumental Variables Approach

Addressing these issues for causal inference requires some exogenous (w.r.t. individual decisions to acquire ICT skills) variation in ICT skills. Since we are concerned about unobserved heterogeneity that determines the *demand* for ICT skills at the individual level, a natural place to look for such variation is the *supply* side, i.e., availability of internet and its costs (Angrist and Krueger (2001)). The dramatic expansion of the cell phone tower in the last few decades in India provide a source of such supply side variation because most of the internet access in 2020 (our survey year) is based on mobile phones. As discussed in detail in section (3) above, cell towers effectively "blanketed" the whole country by 2020 including villages which yet to have connection to electricity grids, and such a vast expansion of cell towers meant that they were not targeted to the rich districts with high growth prospects (see below).

But while district level tower density captures the variation in the supply of telecom services across different districts, a potential threat to the exclusion restriction is that it might be correlated with the economic structure and development potential of a district. The correlation between tower density and district characteristics are likely to be strong when the telecom operators have to build their own towers, and thus they enter into only the high profit areas to recover their fixed costs.[26] As discussed before, in India the expansion of cell towers and cell phone service has been dramatic, effectively blanketing the whole nation by 2020 (Purnell (2018)). This suggests that the enodgeneity that arises from targeting markets with high potential by telecom operators is much less of a concern in India. This, however, does not mean that the spatial variation in tower density across districts is not correlated with economic characteristics at all. We would expect that a tower company

[26]As noted in section (3) above, this in fact is the case in many countries where Towerco industry is not developed.

such as Indus tower to build more towers in a district with more educated (high income) population, and more ICT-intensive economic activities. Such correlation can lead to a direct impact of tower density on the labor market outcomes because tower density acts as a proxy for the omitted economic potential and structure of a district. For example, the share of agricultural employment in a district is likely to be negatively correlated with ICT skills because farming is not dependent on ICT skills in most cases. The labor market gains from ICT skills for an individual will also be lower in a district where most of the employment is in farming.[27]

To address this concern, we adopt a three-step approach. First, we include a vector of district level characteristics including employment structure and education of the population using the 2011 census data. Since our outcome variables and the ICT skills index are measured in 2020, these district level control variables are predetermined in our analysis.[28] Second, and more importantly, we develop a leave-own-out IV (henceforth LOU IV) strategy that exploits a particular aspect of the telecom regulation policy of GOI which created 22 telecom circles in 1994 (discussed in more detail in the next subsection). The LOU IV is constructed for a district by leaving out the district itself, and using other districts in a larger spatial unit, a telecom circle (see the next subsection). Third, we check the sensitivity of our conclusions to local violation of the exclusion restriction by allowing for very small direct impact of the instrument on the outcomes in the IV bounds approach developed by Conley et al. (2012).

(5.2) Telecom Circles and A Leave-own-out (LOU) IV

Under the first National Telecom Policy (NTP) of Government of India (GOI), the country was divided into 22 separate administrative units known as a 'Telecom Circle' in 1994. The telecom circles mostly overlapped with state boundaries except for smaller Northeastern states which were combined to form a single circle. The most populous state, Uttar Pradesh, was divided into two circles, while the smaller union territories were combined with adjacent state circle. The original telecom circles persisted over time, hence

[27] The estimates from a regression of district tower density on different economic characteristics of a district suggests that density is higher in a district with more educated population. But share of agricultural employment has no significant correlation. The details are available from the authors.

[28] If we use the 2001 census for the district level controls, the main conclusions in this paper remain robust. The details are available upon request.

the newly created states of Chattisgarh, Jharkhand, Uttarakhand, and Telenganna remain with the parental state circles.

The Government of India (GOI) auctioned spectrums based on telecom circles, so different circles may have different telecom spectrum and different numbers of operators. The evidence also suggests that the quality (upload and download speeds) and cost of internet access vary across telecom circles. We rely on the fact that within a circle, the quality and policies of telecom operators remain the same creating a correlation between a district tower density and tower density in the circle excluding the concerned district. We construct our instrument in the following way.

$$tdens_{dc} = \beta_0 + \beta tdens_{(-d)c} + \sum_{k=1}^{21} \beta_k \left(\gamma_k \times tdens_{(-d)c}\right) + \sum_{l} \delta_l \left(x_{dl} \times tdens_{(-d)c}\right) + \varepsilon_{dc} \quad (3)$$

where $tdens_{dc}$ is tower density in district d in circle c, $tdens_{(-d)c}$ is the circle-level leave-own-out (LOU) tower density where the concerned district d is excluded when calculating density. We allow the relationship between LOU circle-level tower density and district tower density to vary across circles. $x_{dl} \times tdens_{(-d)c}$ is the interaction of characteristic l in district d with its LOU circle-level tower density. The included district characteristics are district population density, urban share in population, share of agricultural laborers and share of cultivators in the workforce, share of Scheduled Castes and share of Scheduled Tribes in district population, district literacy rate, and share of higher secondary or above in district population (data from 2011 census).[29] We estimate equation (3) using OLS from district level data that consists of 640 districts (district boundaries are according to 2011 census), and predict tower density for the district, $\widehat{tdens}_{dc}$. We use the predicted tower density $\widehat{tdens}_{dc}$ for the districts as an instrument in the 2SLS regressions for estimating the effects of ICT skills on labor market outcomes, after conditioning on all the district level variables used for interactions in equation (3) above along with state-region fixed effects and changes in employment structure in a district (see equation (4) below).

An advantage of this strategy is that we convert an underlying over-identified model (the interactions of $tdens_{(-d)c}$ with circle dummies and district characteristics are additional instruments) into a just-identified empirical model. As emphasized by Angrist and Kolesár (2023) and Kolesár et al. (2015), a just identified model suffers the least bias. Also, Kolesár et al. (2015) point out that this strategy also imposes weaker exclusion restrictions for

[29] The main conclusions of this paper remain intact if we instead use data from 2001 census.

identification because we do not require separate exclusion restrictions on each instruments in the right hand side of equation (3).

We use predicted tower density $\widehat{tdens}_{dc}$ from equation (3) above as an instrument for ICT skills of individuals, and estimate the following outcome equation using the 2SLS estimator.

$$Y_{idrc} = \alpha_0 + \gamma ICT_{idrc} + \delta X_{idrc} + \psi X_{drc} + \pi(\Delta E_{drc}) + \Gamma_{rc} + \varphi_{idrc} \tag{4}$$

As our instrument, $\widehat{tdens}_{dc}$, varies across districts, we replace the district-fixed effects ξ_{drc} in equation (1) earlier with state-region fixed effects Γ_{rc}, and add district-level controls (measured in 2011) used in the prediction of district tower density in equation (3) above.[30] We provide evidence that state-region fixed effects (Γ_{rc}) and district level controls (X_{drc}) together constitute a set of sufficient statistics for district fixed effects in our estimation. We also include a vector of individual level controls X_{idrc} such as education level, age, gender, caste, religion (Muslim dummy), and location (rural dummy). These individual level controls reduce the omitted variables bias; for example, education and gender variables account for the possibility that more educated people and men are more likely to acquire ICT skills and also have better labor market outcomes. Given that our instrument, $\widehat{tdens}_{dc}$, is constructed at the district level, we cluster the standard errors at the district level.

The identification assumption in our IV strategy is that conditional on state-region fixed effects, district level controls used in equation (4) including the changes in employment structure, and individual-level selection on observables as captured by the vector X_{idrc}, the predicted tower density at the district level from equation (3) significantly influences ICT skills in a district, but does not affect the labor market outcomes of an individual directly (exclusion restriction). The exclusion restriction is plausible because the predicted tower density of a district rely on tower density in the relevant telecom circle excluding the concerned district, and thus it should not have any direct effect on outcomes in the concerned district. The state-region fixed effects plus district controls ensure that the tower density of the LOU districts do not capture spatial correlations in economic structure across districts or the effects of cell towers on the employment structure in a district.[31]

[30]Note that there are 83 state-region fixed effects, while there are only 22 telecom circles. Thus we do not need circle fixed effects in the specification. Also, note that the conclusions remain robust if we use 2001 measures of district level controls instead of 2011 measures.

[31]An average district in India according to the 2011 Census has an area of 4948 square Km and a population of 1.89 million.

(6) Evidence

(6.1) Preliminary Estimates

We begin with the OLS estimates in Table 3. We provide estimates for various sets of control variables to gauge the sensitivity of the estimates. Column 1 of Table 3 reports the estimates without any controls or fixed effects. The estimates suggest that having ICT skills increases the probability of labor force participation, employment, salaried employment, and per capita household expenditure, and reduces the probability of daily wage employment, and self employment. The second column includes district fixed effects and no other controls; the estimated effects barely change for most of the outcomes. Even for the outcomes with a moderate change in the magnitude, for example, probability of self-employment and per capita expenditure, the estimates with district fixed effects in column (2) are not significantly different from the corresponding estimates in column (1) at the 5 percent level. This evidence suggests that district level observed and unobserved heterogeneity are not a concern for estimating the causal effects. This has important implications for our research design as the IV estimates using LOU district-level cell tower density as the instrument are unlikely to be biased because of omitted heterogeneity at the district level. As a conservative strategy, we include the district level controls discussed in equation (4) above in all IV regressions.

In contrast, when we include a vector of individual level controls such as education, gender, age, caste, and religion in column (3), the magnitudes of the estimates are substantially smaller for all outcomes except for self-employment, and the differences between columns (2) and (3) are statistically significant at the 5 percent level. This is consistent with the discussion in section (4) above that individual level unobserved heterogeneity including omitted cognitive and noncognitive abilities are important sources of omitted variables bias in our context.

In column (4) of Table 3, we include state-region fixed effects and a rich set of district level controls, including the changes in employment structure, instead of district fixed effects. The motivation for this exercise comes from the fact that our identifying variation for the IV estimates below is at the district level, as the LOU tower density calculated in equation (3) varies across district. So when implementing the IV estimation, we cannot include district fixed effects in the specification. We use district characteristics and state-region fixed effects to capture heterogeneity in economic structure across districts. We control for the changes in the share of services and manufacturing employment in a district from 2004 to 2020 to take into account the firm-side responses to cell-tower expansions in a district. This reflects the consideration that ICT skills are not important for agricultural employment in the context of India.[32] If this strategy is successful, then we should not see any substantial differences in the estimates between columns (3) and (4). The evidence in Table 3 shows that the estimates in columns (3) and (4) do not vary substantially, and in some cases, they are virtually identical, suggesting that the set of district controls together works as a sufficient statistic for district fixed effects.

(6.2) IV Estimates

The IV estimates are reported in Table 4. All the regressions include the individual and district level controls in column (4) in Table 3. The estimated effect of ICT skills index on labor force participation and employment are not significant at the 5 percent level. This is in sharp contrast to the positive and significant effects we found earlier in the OLS estimates. This suggests that the worry about positive selection bias in the OLS estimates is important for these two outcomes.

The IV estimates suggest a positive impact on the probability of regular salaried employment and per capita expenditure, and a negative impact on the likelihood of self-employment and daily wage employment, and the estimates are significant at the 5 percent level. The effect sizes are substantial. A 10 percentile higher ICT skills index increases the probability of regular salaried employment by 6.5 percentage points, and reduces the probability of self employment by 3.6 percentage points and the probability of daily wage employment by 3.4 percentage points. The evidence also suggest a 9.5 percent higher per capita expenditure when the ICT skills index is 10 percentile higher.

The IV estimates are substantially larger than the corresponding OLS estimates in column (4) of Table 3 discussed earlier. While attenuation bias in the ICT skills index is likely to an important factor in explaining the larger IV estimates, a second plausible expansion is that the cell tower expansion affects the ICT skills of subgroups for which the labor market impacts are large.

[32]Since services sector is the most ICT-intensive, we check the robustness of our results by controlling for only the changes in the share of services employment in a district. The details are available from the authors upon request.

Robustness Checks

In this section, we report and discuss evidence on whether the conclusions in Tables 3 are robust in two respects: (i) alternative construction of the LOU IV, and (ii) different age ranges for defining the estimation sample. In section (4.2), we rely on the circle-level LOU tower density and its interactions with district level characteristics for identifying variation. One might wonder whether the main conclusions survive if we do not use the interaction of district characteristics with the LOU tower density for identification, thus setting $\delta_l = 0$ for all l in equation (3) above. The IV estimates from this approach are reported in Panel A of Table 5. The conclusions regarding salaried employment, self-employment, and per capita expenditure remain intact, with larger magnitudes of the effects relative to the estimates in Table 4. In contrast, the estimated effect on the probability of daily wage work becomes insignificant (at the 10 percent level) and is numerically small, even though it bears a negative sign consistent with the estimate in Table 4. However, we note a caveat regarding these estimates in panel A of Table 5: the IV lacks power in this case. Since we convert the empirical model to a just-identified one using equation (3) (with $\delta_l = 0$ for all l), the weak instrument bias may not be as damaging (see Angrist and Kolesár (2023)). With a just-identified model, weak instrument robust Fuller LIML estimates are virtually identical to the 2SLS estimates reported in Table 4.[33]

We also construct another instrument by using the district characteristics from 2001 census in equation (3) in place of 2011 census used earlier for the results in Table 4. This is motivated by the fact that cell phone access in 2001 was much lower compared to 2011 even though most of the telecom expansion occurred after 2010. The IV estimates using this alternative instrument are reported in panel B of Table 5. The conclusions regarding salaried employment, self-employment, and per capita expenditure remain intact, but the estimated effect on daily wage employment is no longer significant at the 10% level and smaller in magnitude.

A last robustness check we implement deals with the question of whether the conclusions are sensitive to small changes in the age range used to define the estimation sample. We estimate the effects of ICT skills for a number of different age ranges, and the evidence suggests that the conclusions reached based on Table 4 are robust to small changes in the

[33]Estimates from the Fuller LIML estimator are available upon request.

age ranges. For the sake of brevity, we report the estimates for two age ranges: 25-55 years (panel C of Table 5) and 20-50 years (panel D of Table 5).

(6.3) Relaxing the Exclusion Restrictions: Evidence from Conley et al. (2012) Bounds

The point estimates from the 2SLS estimator in Table 4 are valid under the strict exclusion restriction that our instrument, predicted tower density in a district based on the LOU circle-level tower density, has exactly zero direct impact on the outcomes of interest. In this section, we report evidence on the question whether the main conclusions from Table 4 remain valid if we allow for very small direct impact of the instrument on the outcomes. To this end, we implement the IV bounds approach developed by Conley et al. (2012). To see intuition of their approach in our context, consider the following extension of the empirical model in equation (4) (use j as the outcome index):

$$Y_{jidrc} = \alpha_0 + \gamma ICT_{idrc} + \delta X_{idrc} + \psi X_{drc} + \pi(\Delta E_{drc}) + \Gamma_{rc} + \rho_j \widehat{tdens}_{dc} + \varphi_{idrc} \quad (5)$$

The identification assumption for the IV estimates in Table 4 is $\rho_j = 0$. We do not have any prior regarding the sign or magnitude of the direct effect ρ_j which captures some unspecified complex channel through which LOU tower density can affect outcome j. We thus follow the suggestion of Conley et al. (2012) and estimate bounds on the confidence interval under the assumption that $\rho_j \in [-\eta, +\eta]$ where $\eta > 0$ is arbitrarily small.

To fix a scale for the value of η, we first estimate the reduced form effect of the instrument on an outcome, i.e., estimate equation (5) excluding the endogenous variable ICT_{idrc}. Denote the estimated reduced form impact of the instrument on outcome Y_j by $\hat{\Phi}_j$. Then we set η_j equal to 1%, 3%, 5% and 10% of the reduced form impact, and estimate the Conley et al. (2012) bounds for $\eta_j = 0.01\hat{\Phi}_j$; $\eta_j = 0.03\hat{\Phi}_j$; $\eta_j = 0.05\hat{\Phi}_j$; and $\eta_j = 0.10\hat{\Phi}_j$. We implement the union of confidence interval (UCI) approach that does not impose any distributional assumptions and give the most conservative bounds. The estimated lower and upper bounds of the 90% confidence intervals for our four outcome variables are reported in Table 6. The bounds do not include zero in any of the estimated effects where we found a significant causal impact of ICT skills in Table 4. This is reassuring because it addresses the concern that our conclusions might be invalidated by small departures from the exclusion restriction imposed in the IV estimates in Table 4.

(7) Heterogeneity: The Role of Gender, Caste, Religion, and Location

The results discussed so far refer to the effects for the whole sample. We now explore whether the effects vary systematically with gender, caste, religion, and location of an individual. For this exercise, we focus on the outcomes for which the evidence in Table 4 suggest a significant effect of ICT skills.

There are two reasons to expect that the effects might be different across different subgroups. First, since we rely on an IV strategy for identification, the estimates capture the effects for a subset of the population whose treatment status (or treatment intensity) is affected by the variations in our instrument (i.e., the compliers subset). The heterogeneity can arise from the fact that the subset of compliers are different across different subgroups. Second, different subgroups may have different economic characteristics which partly determine the effects of ICT skills. For example, if a higher caste person has better labor market network for salaried employment in the formal industry and public sector, then we expect the causal effects of ICT skills on the likelihood of salaried employment to be larger in magnitude for high caste subgroup (compared to the low caste). In contrast, lower caste people may have a higher probability of getting a public sector job once they acquire ICT skills because of employment quota targeted to the Scheduled Castes (SC) and Scheduled Tribes (ST) by the Government of India.[34] If this quota effect is strong enough to more than offset the adverse network effect, the causal effects of ICT skills on labor market outcomes are likely to be higher for SC and ST.[35] We also analyze if the effects are different for the younger generation who grew up with the expansion of the cell towers and cell phones in India. Having ICT skills may make a bigger difference for the older people because most of them do not have ICT skills.

[34]Under the employment quota system, 15 percent and 7.5 percent of government posts are reserved for SC and ST respectively. The reservation policy also provides relaxations in age limit and exemption from application/exam fees. At any point these percentages of jobs need to be held by candidates appointed by reservation. Whenever their representation comes down, it would be completed. If the earmarked posts are not filled because of the non-availability of suitable candidates, those posts are marked as backlog vacancies and carried forward to the subsequent recruitment year. Evidence suggest that many of the reserved posts remain unfulfilled because of non-availability of SC/ST candidates. As per the data from the Government of Indiaâs Department of Personnel & Training (DoPT), the backlog vacancies in SC and ST category was 8,223 and 6,955 in the year 2016 and it increased to 14,366 and 12,612 in the year 2019 (source: https://www.newindianexpress.com/nation/2021/Oct/08/backlog-of-scst-list-vacancies-in central-ministries-doubles-in-4-years-2369053.html).

[35]Moreover, better CT skills for SC/ST may reduce the impact of a lack of labor market network through job search channel.

To explore potential heterogeneity in the effects of ICT skills across different subgroups, we estimate the following equation by 2SLS:

$$Y_{idrc} = \alpha_o + \gamma ICT_{idrc} + \alpha_g D_g + \gamma_g \left(ICT_{idrc} \times D_g\right) + \delta X_{idrc} + \psi X_{drc} + \pi(\Delta E_{drc}) + \Gamma_{rc} + \varphi_{idrc} \quad (6)$$

where D_g is dummy for group $g = f, R, M, LC, y$, with $f = female$, $R = rural$, $M = Muslim$, $LC = Lower\ Caste$ and $y = young$. For example, we have $D_f = 1$ for women and zero for men when the focus is on understanding the gender heterogeneity. Estimating equation (6) allows for differences in both the intercept and the slope across relevant subgroups. Our focus is on the parameter γ_g that provides an estimate of the difference in the causal effects between the groups. For example, a $\gamma_f < 0$ when estimating the effects of ICT skills on salaried employment would suggest a gender penalty against women in highly paid and stable salaried employment. Note that equation (5) has two endogenous variables: ICT_{idrc} and $ICT_{idrc} \times D_g$. We use the interaction of LOU tower density with the group dummy, i.e., $\widehat{tdens}_{dc} \times D_g$ as a second instrument.

The estimates are reported in Table 7: panel A for gender, B for caste, C for religion, D for rural-urban, and E for different age groups. The evidence suggests that the instrument based on LOU circle-level tower density has adequate power to estimate the effects on the outcome variables. Since we have more than one endogenous variables in the IV model (5), to judge the power of the instruments, we rely on the Sanderson and Windmeijer (2016) weak IV tests for multiple endogenous variables.[36]

(7.1) Gender: Is There a Women Penalty?

Panel A of Table 7 reports the estimates for potential gender differences. The evidence suggests that there is a gender penalty against women in the labor market: the positive impact of ICT skills on salaried employment is smaller for women, a 10 percentile higher ICT skills index implies a 4 percentage points higher probability of salaried employment for women, while the corresponding estimate for men is more than double at 8.4 percentage points. Similarly, the estimates for self-employment and daily wage employment also indicate substantially lower magnitudes of the impacts for women. When considered together with the fact that women lag far behind in acquiring ICT skills, the evidence suggests that women suffer double disadvantages in making transition to salaried employment.[37]

[36]As discussed by Andrews et al. (2019), the standard Kleibergen-Paap F statistic is suitable for a single endogenous variable.

[37]For example, 29.3 percent of men can copy or move a file or folder, it is only 16.2 percent for women in

The estimates in the last column of panel A of Table 7 suggest that a household with a woman who have ICT skills has substantially higher per capita expenditure. This may seem counterintuitive given gender bias in India. However, this higher estimated effect on per capita expenditure for women reflects two factors: (i) the proportion of women with high ICT skills is much lower, and (ii) assortative matching in marriage implies that the husband (or partner) of a woman with high ICT skills is likely to have high ICT skills, thus making the family stock of ICT skills much higher. Note that there is a substantial proportion of men with ICT skills who cannot find a bride with ICT skills because of relative scarcity of ICT skills among women. Thus, on an average, the family ICT skills (husband + wife) are lower for the subsample of men with high ICT skills when compared to the subsample of women with high ICT skills.

(7.2) Caste and Labor Market: Are the Returns to ICT Skills Lower for the Lower Castes?

Panel B of Table 7 contains the estimates for lower vs. upper castes. We estimate the effects separately for two groups of lower castes: Scheduled Castes (SC) and Scheduled Tribes (ST). Note that there are three endogenous variables in this empirical model, as we have two caste interactions with the ICT skills index. We thus use three instruments, with the interactions of LOU tower density with SC and ST dummies as additional instruments. The Sanderson-Windmeijer F statistics suggest that the IVs have reasonable power for all three endogenous variables.

The estimates suggest that ICT skills have a larger positive impact on the probability of salaried employment for both SC and ST (relative to the other social groups including the higher castes), and the magnitude of the effect is larger for the SC. Higher ICT skills lead to a movement away from daily wage employment to salaried employment for the SC, while for ST, the occupational transition occurs from self-employment to salaried employment. The finding that the SC and ST enjoy higher access to salaried employment compared to the other social groups might seem surprising, given the accumulated evidence that these lower castes face lower economic opportunities and labor market disadvantages. However, as we noted before, this can be explained by the employment quotas for SC and ST in public sector employment. Higher ICT skills increases the probability of regular salaried job

our sample.

relatively more for SC/ST because of relatively more abundant government jobs for SC/ST.

The estimated effect of the interaction SC/ST dummies with ICT skills index on per capita household expenditure in the last is small in magnitude and statistically not significant. This suggests that they get similar benefits from a better access to employment opportunities to the other social groups including the higher castes.

(7.3) The Role of Religion: Are the Muslims Left Behind?

Panel C of Table 7 reports the estimated effects of ICT skills on employment outcomes and per capita household expenditure for Muslim vs. non-Muslim population. There are no significant differences between Muslim and non-Muslim population in the impact on salaried employment. However, there are differences in the effects on the employment pattern: the probability of daily wage employment declines much more substantially for Muslims with ICT skills, but the decline in the probability of self-employment is much smaller for Muslims compared to the other social groups. This pattern suggests that the Muslims with ICT skills move from daily wage employment to self-employment (see columns 2 and 3 in panel C of Table 7).

It is difficult to judge whether a transition from daily wage employment to self-employment is an improvement without looking at the impacts on household per capita expenditure. The estimates in the last column suggests a small positive coefficient, but it is not statistically significant at the 10 percent level. Thus, the gains in per capita expenditure from ICT skills are broadly similar in magnitude across Muslim and non-Muslim population. This suggests that acquiring ICT skills might be an important avenue for economic improvements for Muslims in India.

(7.4) Rural vs. Urban: Does Location Matter?

The evidence in panel D of Table 7 suggests that the effects of ICT skills on labor market outcomes and household per capita expenditure do not vary substantially between rural and urban areas. The estimated coefficient of the rural dummy interacted with ICT skills index is small and statistically not significant at the 10 percent level for salaried employment, daily wage employment, and per capita expenditure, even though point estimates have positive signs for salaried employment and per capita expenditure. The interaction effect is negative for self-employment, possibly capturing movements away from farming which is the most common form of self-employment in rural India.

(7.5) Young vs. Old: Do ICT Skills Matter More for the Older Generation?

The last panel of Table 7 contains the estimates for young (21-40 years) vs. old (41-60) age cohorts. The 21-40 years old people in the survey year 2020 were 1-10 years old in 2000, and they grew up with the expansion of cell towers and cell phones in the decades of 2000s and 2020s. We thus expect that many in this generation were able to acquire ICT skills compared to the older generation. For example, among the younger people (21-40 years old in 2020), 38.1% male and 15.8% female could copy or move a file or folder, the corresponding numbers for the older generation (41-60 years) are 22.7% and 6.1% respectively. This suggests that supply of ICT skills among the older generation is limited, and this might imply that returns to ICT skills in the labor market are substantially higher for this group.

The estimates suggest that there are no substantial differences across the age groups in the effects on the probability of salaried employment and daily wage employment, but the decline in self-employment caused by ICT skills is substantially smaller in magnitude (about half) for the younger cohorts. There is a positive impact on the per capita expenditure for the younger cohort, but the magnitude is about half of that for the older age cohorts. The larger impact on the older cohorts might capture higher wage returns for them as the effects on the probability of getting salaried employment or daily wage employment do not vary substantially across age groups.[38] The higher wage returns for the older cohorts is likely to capture life-cycle effects.

(8) Mechanisms: Are Education and Job Search Channels Important?

As discussed before, there are three major channels through which ICT skills can affect the labor market outcomes: (i) ICT skills improve employability in higher paid and stable salaried employment in public sector, and also private industries and service sectors; (ii) ICT skills help in job search, widening the radius of labor market; and (iii) ICT skills may improve educational outcome. In this section, we provide evidence on the importance of the job search and education channels in our results.

We have data on education level, and we control for education of an individual in our main estimates in Table 4. This means that if education is an important mechanism, our estimates in Table 4 do not capture the effects of ICT skills mediated through this channel.

[38] We do not have the wage data to explore this in depth.

To check whether education is, in fact, an important channel, we drop the education dummies from the empirical specification and reestimate the effects of ICT skills (see panel A in Table 8). When compared to the estimates in Table 4, the estimates in Table 8 are not substantially different. This suggests that education might not be a major channel through which the effects of ICT skills work in our data. This conclusion is also supported by the evidence discussed in the preceding section that suggests that there are no substantial differences in the effects between the younger and older cohorts. If ICT skills are important for educational outcomes, this effect will be found in the younger cohorts who did not already finish schooling before the expansion of cell towers and internet in India.

The survey data we use do not contain information on job search. But we have information on whether the father is engaged in salaried employment. We construct a sample of individuals whose fathers reside in the household and aged 60 and less. From this sample, we know the labor market status of an individualâs father. When an individual's father is in the salaried employment, he/she has access to information about job qualifications and vacancies through father's labor market network. They do not rely on the job openings posted online because they get the information early on even before the vacancy is advertised. In contrast, for a job seeker without such family information network (father not in salaried employment), online job search will be important. We check whether the impact of ICT skills on salaried employment is larger for the children of fathers not in salaried employment, as implied by the job search channel. The interaction of a dummy for father not in salaried employment with the ICT skills index is positive and significant at the 5 percent level (see panel B in Table 8). ICT skills lead to a 2.3 percentage points higher probability for children of fathers who are not in salaried employment (compared to the children with fathers in salaried employment). This evidence supports the job search channel. Note that the children of fathers in salaried employment not only enjoy access to information about jobs (low cost of job search), they are also likely to have higher bargaining power because of father's network (it is more likely to get a job if the manager hiring is a friend of your father). Thus, the estimated effect of job search channel in Table 8 might be biased downward.

(9) Conclusions

Using the first nationally representative large data set on ICT skills at the individual level in India (Multiple Indicators Survey, 2020), we provide evidence on the effects of ICT skills on labor market outcomes and household per capita expenditure. We analyze the impact of ICT skills both at the extensive margins (labor force participation and employment) and intensive margins (transitions among different types of employment) of the labor market. The main challenge in identifying the effects is unobserved individual heterogeneity that affects both ICT skills acquisition and labor market outcomes. To deal with this, we develop an instrumental variables strategy that leverages the dramatic expansion of cell towers in the 2010s spearheaded by Towercos and Reliance Jio phone as a source of supply shock, and develops a leave-own-out (LOU) instrument exploiting an institutional feature of Indian telecom regulation: the "telecom circles" which were used by the government to auction telecom spectrum.

The estimates from the IV approach suggests that there is no perceptible impact of ICT skills at the extensive margins of the labor market: the effect is small in magnitude and not significant at the 10 percent level for both labor force participation and unemployment. In contrast, higher ICT skills increase the probability of salaried employment substantially, and lead to a higher per capita household expenditure, suggesting a higher welfare. The evidence indicates employment transitions at the intensive margins, away from daily wage employment and self-employment for individuals with a higher ICT skills. These conclusions are robust to local violation of the exclusion restriction imposed in the IV estimation.

We find evidence of substantial heterogeneity in the effects of ICT skills. Women face a gender penalty in the sense that the positive impact on salaried employment is about 30 percent smaller. In contrast, the impact of ICT skills on per capita expenditure is higher for women, reflecting the fact that the family stock of ICT skills is much larger in those households where women have higher ICT skills because their husbands also tend to have higher ICT skills due to assortative matching. The lower caste individuals (Scheduled Caste and Scheduled Tribes) with ICT skills enjoy a higher probability of getting salaried employment compared to the rest of the population including the higher castes. We, however, find that Muslims enjoy comparable returns in terms of salaried employment and per capita expenditure once they have ICT skills. There are no significant differences between rural and urban areas. We explore the role of education and job search as potential mediating mechanisms, and find suggestive evidence of a job search channel, but no evidence for any impact through educational attainment as a mediating channel.

References

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American economic review*, 98(3):808–842.

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2019). Democracy does cause growth. *Journal of political economy*, 127(1):47–100.

Aggarwal, A. (2021). Global framework on core skills for life and work in the 21st century. Technical report, International Labor Organisation.

Andersen, T. B., Bentzen, J., Dalgaard, C.-J., and Selaya, P. (2012). Lightning, IT diffusion, and economic growth across US states. *Review of Economics and Statistics*, 94(4):903–924.

Andrews, I., Stock, J. H., and Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11:727–753.

Angrist, J. and Kolesár, M. (2023). One instrument to rule them all: The bias and coverage of just-id IV. *Journal of Econometrics*.

Angrist, J. D. and Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4):69–85.

Bajpai, N. and Biberman, J. (2019). The future of work in india: Adapting to the fourth industrial revolution. Technical report, Earth Institute, Columbia University.

Betz, T., Cook, S. J., and Hollenbach, F. M. (2018). On the Use and Abuse of Spatial Instruments. *Political Analysis*, 26(4):474479.

Caria, S., Orkin, K., Garlick, R., Singh, N., Heath, R., and Andrews, A. (2024). Barriers to search and hiring in urban labour markets. Technical report, Technical Report, Vox Dev Literature 2024.

Carranza, E. and McKenzie, D. (2024). Job training and job search assistance policies in developing countries. *Journal of Economic Perspectives*, 38(1):221–244.

Chiplunkar, G. and Goldberg, P. K. (2022). The employment effects of mobile internet in developing countries. Technical report, National Bureau of Economic Research.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly Exogenous. *The Review of Economics and Statistics*, 94(1):260–272.

Falck, O., Heimisch-Roecker, A., and Wiederhold, S. (2021). Returns to ICT skills. *Research policy*, 50(7):1040–64.

Fruehwirth, J. C., Iyer, S., and Zhang, A. (2019). Religion and Depression in Adolescence. *Journal of Political Economy*, 127(3):1178–1209.

Government of India (2023). Multiple indicators survey in india. Technical report, National Sample Survey Office.

Gupta, A., Kakar, K., Kapoor, G., Srivastava, S., and Mahajan, S. (2020). From Evolution to Revolution: Advancing a Decade of Innovation in Indian Towerco Industry. Technical report, Earnest And Young.

Hjort, J. and Poulsen, J. (2019). The arrival of fast internet and employment in africa. *American Economic Review*, 109(3):1032–1079.

Hjort, J. and Tian, L. (2024). The economic impact of internet connectivity in developing countries. Working paper, University College London.

Houngbonon, G. V., Rossotto, C. M., and Strusani, D. (2021). Enabling A Competitive Mobile Sector in Emerging Markets Through the Development of Tower Companies. Technical report, International Finance Corporation, Washington, DC.

Kolesár, M., Chetty, R., Friedman, J., Glaeser, E., and Imbens, G. (2015). Identification and Inference With Many Invalid Instruments. *Journal of Business & Economic Statistics*, 33(4):474–484.

Mangla, A. and Singh, M. (2021). Study on the Growth of Indian Telecom Sector: Evidence from Post-Liberalization Period. *Orissa Journal of Commerce*, 42(4):130–145.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Non, M., Dinkova, M., Dahmen, B., et al. (2021). Skill up or get left behind?: Digital skills and labor market outcomes in the netherlands. Technical report, CPB Netherlands Bureau for Economic Policy Analysis.

Persson, T. and Tabellini, G. (2009). Democratic capital: The nexus of political and economic change. *American Economic Journal: Macroeconomics*, 1(2):88–126.

Purnell, N. (2018). Two years ago, India lacked fast, cheap internet one billionaire changed all that. *The Wall Street Journal*.

TRAI and NCAER (2012). Telecom Sector in India: A Decadal Profile. Technical report, Telecom Regulatory Authority of India.

Zuo, G. W. (2021). Wired and Hired: Employment Effects of Subsidized Broadband Internet for Low-Income Americans. *American Economic Journal: Economic Policy*, 13(3):447–82.

Table 1: Descriptive statistics, outcome, and control variables

Number of observations	**622,764**
Outcomes:	
Labor Force Participation (1/0)	0.618 (0.486)
Employed (1/0)	0.585 (0.493)
Regular salaried (1/0)	0.130 (0.336)
Self-employment (1/0)	0.269 (0.443)
Daily wage worker (1/0)	0.175 (0.380)
Log of per capita expenditure	9.155 (0.555)
ICT Skills Index	0.000 (3.937)
Instrument Variable	
Predicted tower density per 10000 population in a district based on the tower density in other districts in a telecom circle	22.097 (31.373)
Controls:	
Individual controls	
Rural (1/0)	0.688 (0.463)
Female (1/0)	0.499 (0.500)
Highest level of education	
Below primary (1/0)	198,529 (31.9%)
Primary/middle (1/0)	179,790 (28.9%)
Secondary (1/0)	89,689 (14.4%)
Higher Secondary (1/0)	67,584 (10.9%)
Graduate (1/0)	68,887 (11.1%)
Postgraduate (1/0)	18,285 (2.9%)
Age (indicators for year wise age)	
21	12,166 (2.0%)
22	20,345 (3.3%)
...omitted	
59	4,803 (0.8%)
60	20,547 (3.3%)
Married (1/0)	0.825 (0.380)
Social Group	
Scheduled Tribes (1/0)	57,729 (9.3%)
Scheduled Castes (1/0)	128,757 (20.7%)
Other Backward Castes (1/0)	278,878 (44.8%)
General (1/0)	157,400 (25.3%)
Muslim (1/0)	0.110 (0.313)

Table 1 Cont.

District-level controls:	
District population density, Census 2011	1,315.570 (3,693.774)
Urban population share, Census 2011	0.320 (0.235)
Share of Scheduled Castes in population, Census 2011	0.167 (0.077)
Share of Scheduled Tribes in population, Census 2011	0.087 (0.154)
Share of Agriculture laborers, Census 2011	0.234 (0.139)
Share of cultivators, Census 2011	0.265 (0.151)
District literacy rate, Census 2011	0.637 (0.102)
Higher secondary or above share, Census 2011	0.131 (0.066)
Change in share of service and manufacturing in workforce over 2004-2020, NSS	0.170 (0.119)

Note: The sample is restricted to ages 21-60. Weighted means are reported. Standard deviations reported in parentheses. NSS implies data source is National Sample Survey.

Table 2: Information, communication, and technology (ICT) skills

ICT Skills Index (normalized)*3.973	**0.000 (3.937)**
a. Able to copy or move a file or folder+	0.227 (0.419)
b. Able to use the copy and paste tools to duplicate or move information within a document+	0.215 (0.410)
c. Able to send e-mails with attached files (e.g. document, pictures, videos) +	0.142 (0.349)
d. Able to use basic arithmetic formulae in a spreadsheet+	0.056 (0.230)
e. whether able to connect and install new devices (e.g. modem, camera, printer) +	0.073 (0.259)
f. Able to find, download, install and configure software+	0.118 (0.323)
g. Able to create electronic presentations with presentation software (including text, images, sound, video, or charts) +	0.049 (0.217)
h. Able to transfer files between a computer and other devices+	0.106 (0.308)
i. Able to write a computer program using a specialized programming language+	0.014 (0.119)
Observations	**622,764**

Note: +implies indicator (1/0) variable. The sample is restricted to ages 21-60. Weighted means are reported. Standard deviations reported in parentheses. ICT skill index is derived by the principal component analysis (PCA), and the normalized PCA ICT skill index is multiplied by 3.973. Hence one unit change in ICT skill index means 0.254 Standard Deviation change, i.e., 10 percentiles change in skill distribution.

Table 3: Impact of ICT skills, OLS estimates

	(1)	(2)	(3)	(4)
Labor Force Participation (1/0)	0.020***	0.020***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
R-Square	0.027	0.059	0.540	0.529
Employed (1/0)	0.011***	0.010***	0.004***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
R-Square	0.010	0.047	0.496	0.484
Regular salaried (1/0)	0.027***	0.026***	0.018***	0.018***
	(0.001)	(0.001)	(0.001)	(0.001)
R-Squared	0.150	0.171	0.217	0.210
Self-employed (1/0)	-0.005***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
R-Square	0.019	0.064	0.205	0.190
Daily wage worker (1/0)	-0.012***	-0.013***	-0.010***	-0.009***
	(0.000)	(0.001)	(0.000)	(0.000)
R-Square	0.033	0.073	0.194	0.180
Log per capita expenditure	0.030***	0.025***	0.014***	0.015***
	(0.002)	(0.002)	(0.001)	(0.002)
R-Square	0.158	0.320	0.362	0.316
Observations	622,764	622,764	622,764	622,764
Controls				
District-Fixed effects	NO	YES	YES	NO
State-region Fixed effects	NO	NO	NO	YES
Individual characteristics	NO	NO	YES	YES
District characteristics	NO	NO	NO	YES

Note: The table reports the coefficient on ICT skill index from separate regressions with different outcomes. A one unit change in ICT skill index means 0.254 Standard Deviation change, i.e., 10 percentiles change in skill distribution. All models control gender and rural areas. Standard errors clustered at district-level are reported in parentheses. Individual and district-level controls are reported in Table 1.

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Impact of ICT skills, Instrument variable (IV) estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Labor Force Participation (1/0)	Employed (1/0)	Regular salaried (1/0)	Self-employed (1/0)	Daily Wage work (1/0)	log (per capita expenditure)
ICT Skills Index	-0.010	0.003	0.065***	-0.036*	-0.034**	0.095**
	(0.015)	(0.015)	(0.017)	(0.020)	(0.015)	(0.048)
Observations	622,764	622,764	622,764	622,764	622,764	622,764
Kleibergen-Paap rk Wald F	13.80	13.80	13.80	13.80	13.80	13.80

Notes: (1) The table reports the coefficient on ICT skill index from separate instrument variable (IV) regressions with different outcomes. A one-unit change in ICT skill index means 0.254 Standard Deviation change, i.e., 10 percentiles change in skill distribution. (2) The instrument used for the ICT skill index is district-level predicted cell tower density per 10,000 population. The predicted district tower density is derived from a district-level OLS regression that regresses district tower density on the tower density in the telecom circle leaving the district itself out (leave-own-out) while allowing for the relationship to vary across telecom circles and 2011 Census district characteristics. See Section 5.2 for more details about the IV. (3) All the IV regressions control for the 2011 Census district characteristics that are used for the instrument construction. All models control for state-region fixed effects, individual and district-level controls reported in Table 1. (4) Standard errors clustered at district-level are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Robustness to alternative IV construction and samples

	(1)	(2)	(3)	(4)	(5)	(6)
	Labor Force Participation (1/0)	Employed (1/0)	Regular salaried (1/0)	Self-employed (1/0)	Daily Wage work (1/0)	log (per capita expenditure)
Panel A: Instrument: Predicted tower density based on circle LOU with no district characteristics interaction terms						
ICT Skills Index	0.002	0.008	0.072**	-0.071*	-0.001	0.208**
	(0.024)	(0.025)	(0.030)	(0.040)	(0.022)	(0.084)
Kleibergen-Paap rk Wald F	4.441	4.441	4.441	4.441	4.441	4.441
Panel B: Instrument: Predicted tower density with interaction with census 2001 characteristics, controls census 2001 district characteristics						
ICT Skills Index	-0.010	-0.000	0.066***	-0.057***	-0.017	0.142***
	(0.016)	(0.017)	(0.017)	(0.022)	(0.015)	(0.042)
Kleibergen-Paap rk Wald F	14.24	14.24	14.24	14.24	14.24	14.24
Panel C: Sample, 25-55 age group						
ICT Skills Index	-0.003	0.003	0.065***	-0.039**	-0.031**	0.090**
	(0.015)	(0.015)	(0.017)	(0.020)	(0.015)	(0.045)
Kleibergen-Paap rk Wald F	13.16	13.16	13.16	13.16	13.16	13.16
Panel D: Sample, 20-50 age group						
ICT Skills Index	-0.006	0.008	0.069***	-0.035*	-0.032**	0.085*
	(0.015)	(0.015)	(0.018)	(0.019)	(0.015)	(0.044)
Kleibergen-Paap rk Wald F	13.28	13.28	13.28	13.28	13.28	13.28

Notes: (1) The table reports the coefficient on ICT skill index from separate instrument variable (IV) regressions with different outcomes. A one-unit change in ICT skill index means 0.254 Standard Deviation change, i.e., 10 percentiles change in skill distribution. (2) All models control for state-region fixed effects, individual and district-level controls reported in Table 1. (3) In Panel A, the instrument used for the ICT skill index is district-level predicted cell tower density per 10,000 population where the predicted district tower density is derived from a district-level OLS regression that regresses district tower density on tower density in telecom circle leaving the district out while allowing for the relationship to vary across telecom circles. (4) In Panel B, the instrument, predicted district tower density, is derived a district-level OLS regression that regresses district tower density on tower density in telecom circle leaving the district out while allowing for the relationship to vary across telecom circles and 2001 Census district characteristics. All the IV regressions in Panel B control for the 2001 Census district characteristics that are used for the instrument construction. (5) In Panel C and D, the instrument used is predicted tower density that is derived from a district-level OLS regression that regresses district tower density on tower density in telecom circle leaving the district out while allowing for the relationship to vary across telecom circles and 2011 Census district characteristics. See Section 5.2 for more details about the IV. (6) Standard errors clustered at district-level are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Relaxing the Exclusion Restriction: Conley et al. Bound Estimates

		Lower Bound	Upper Bound
Regular salaried (1/0)	$\rho \in [-0.01 * \hat{\Phi}, 0.01 * \hat{\Phi}]$	0.036	0.094
	$\rho \in [-0.03 * \hat{\Phi}, 0.03 * \hat{\Phi}]$	0.035	0.095
	$\rho \in [-0.05 * \hat{\Phi}, 0.05 * \hat{\Phi}]$	0.035	0.096
	$\rho \in [-0.1 * \hat{\Phi}, 0.1 * \hat{\Phi}]$	0.033	0.099
Self-employed (1/0)	$\rho \in [-0.01 * \hat{\Phi}, 0.01 * \hat{\Phi}]$	-0.068	-0.004
	$\rho \in [-0.03 * \hat{\Phi}, 0.03 * \hat{\Phi}]$	-0.068	-0.003
	$\rho \in [-0.05 * \hat{\Phi}, 0.05 * \hat{\Phi}]$	-0.069	-0.003
	$\rho \in [-0.1 * \hat{\Phi}, 0.1 * \hat{\Phi}]$	-0.070	-0.002
Daily Wage work (1/0)	$\rho \in [-0.01 * \hat{\Phi}, 0.01 * \hat{\Phi}]$	-0.059	-0.009
	$\rho \in [-0.03 * \hat{\Phi}, 0.03 * \hat{\Phi}]$	-0.059	-0.009
	$\rho \in [-0.05 * \hat{\Phi}, 0.05 * \hat{\Phi}]$	-0.060	-0.008
	$\rho \in [-0.1 * \hat{\Phi}, 0.1 * \hat{\Phi}]$	-0.061	-0.007
log (per capita expenditure)	$\rho \in [-0.01 * \hat{\Phi}, 0.01 * \hat{\Phi}]$	0.015	0.174
	$\rho \in [-0.03 * \hat{\Phi}, 0.03 * \hat{\Phi}]$	0.014	0.175
	$\rho \in [-0.05 * \hat{\Phi}, 0.05 * \hat{\Phi}]$	0.012	0.176
	$\rho \in [-0.1 * \hat{\Phi}, 0.1 * \hat{\Phi}]$	0.009	0.179

Notes: (1) This table reports the 90 percent confidence interval bounds for the estimates of the coefficient on ICT skill index under the assumption that the instrumental variable is "plausibly exogenous" in the sense that the instrument can have arbitrarily small direct impact on the outcome (Conley et al. 2012). (2) All models control for state-region fixed effects, individual and district-level controls reported in Table 1. Standard errors are clustered at district-level. (3) The instrument, predicted cell tower density, is derived from a district-level OLS regression that regresses district tower density on tower density in telecom circle leaving the district out while allowing for the relationship to vary across telecom circles and 2011 Census district characteristics. (4) All the IV regressions control for the 2011 Census district characteristics that are used for the instrument construction. (5) $\hat{\Phi}$ is the reduced form impact of the instrument on the outcome variable, ρ is the hypothetical direct impact of the instrument in an IV model. The lower and upper bounds are the estimated effects of ICT skills on the relevant outcome of interest given that ρ belongs to a specified interval. See Section 6.3 for more details.

Table 7: Heterogeneity Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Labor Force Participation (1/0)	Employed (1/0)	Regular salaried (1/0)	Self-employed (1/0)	Daily Wage work (1/0)	log (per capita expenditure)
Panel A: Gender (omitted: Men)						
ICT Skills index	-0.015	-0.003	0.084***	-0.046***	-0.050***	0.074*
	(0.013)	(0.013)	(0.016)	(0.017)	(0.013)	(0.041)
Female*ICT Skills Index	0.011	0.014*	-0.044***	0.022**	0.036***	0.048***
	(0.008)	(0.007)	(0.008)	(0.009)	(0.006)	(0.016)
First Stage F Statistics						
Sanderson and Windmeijer F, 1	19.14	19.14	19.14	19.14	19.14	19.14
Sanderson and Windmeijer F, 2	19.76	19.76	19.76	19.76	19.76	19.76
Panel B: Social Groups (omitted: Non-SC/ST)						
ICT index	-0.011	0.002	0.060***	-0.038**	-0.028**	0.095**
	(0.015)	(0.014)	(0.016)	(0.019)	(0.014)	(0.045)
Scheduled Tribe*ICT Skills Index	0.001	0.002	0.028**	-0.035*	0.008	-0.002
	(0.016)	(0.018)	(0.012)	(0.018)	(0.024)	(0.039)
Scheduled Caste*ICT Skills Index	0.006	0.003	0.035***	0.017*	-0.052***	0.018
	(0.008)	(0.009)	(0.009)	(0.010)	(0.010)	(0.024)
First Stage F Statistics						
Sanderson and Windmeijer F, 1	15.67	15.67	15.67	15.67	15.67	15.67
Sanderson and Windmeijer F, 2	37.60	37.60	37.60	37.60	37.60	37.60
Sanderson and Windmeijer F, 3	46.27	46.27	46.27	46.27	46.27	46.27
Panel C: Religion (omitted: non-Muslims)						
ICT index	-0.010	0.004	0.065***	-0.036*	-0.033**	0.094**
	(0.015)	(0.015)	(0.017)	(0.019)	(0.015)	(0.047)
Muslim*ICT Skills Index	-0.013	-0.012	-0.011	0.030***	-0.036***	0.026
	(0.009)	(0.009)	(0.012)	(0.012)	(0.012)	(0.023)
First Stage F Statistics						
Sanderson and Windmeijer F, 1	13.72	13.72	13.72	13.72	13.72	13.72
Sanderson and Windmeijer F, 2	24.05	24.05	24.05	24.05	24.05	24.05
Panel D: Area (omitted: Urban)						
ICT index	-0.010	0.003	0.064***	-0.035*	-0.034**	0.094**
	(0.015)	(0.015)	(0.018)	(0.019)	(0.016)	(0.048)
Rural*ICT Skills Index	-0.013	-0.020	0.017	-0.026	-0.015	0.023
	(0.013)	(0.014)	(0.012)	(0.018)	(0.011)	(0.040)
First Stage F Statistics						
Sanderson and Windmeijer F, 1	11	11	11	11	11	11
Sanderson and Windmeijer F, 2	11.47	11.47	11.47	11.47	11.47	11.47

Panel E: Age Group (omitted: Age 41-60)						
ICT index	-0.034	-0.016	0.070***	-0.064**	-0.035*	0.140**
	(0.022)	(0.022)	(0.022)	(0.027)	(0.020)	(0.064)
Age 21-40* ICT Skills Index	0.032***	0.023**	-0.011	0.039***	0.003	-0.069***
	(0.010)	(0.009)	(0.009)	(0.011)	(0.008)	(0.026)
First Stage F Statistics						
Sanderson and Windmeijer F, 1	19.44	19.44	19.44	19.44	19.44	19.44
Sanderson and Windmeijer F, 2	34.49	34.49	34.49	34.49	34.49	34.49
Observations	622,764	622,764	622,764	622,764	622,764	622,764

Note: The table reports the coefficient on ICT skill index from separate instrument variable (IV) regressions with different outcomes. A one unit change in ICT skill index means 0.254 Standard Deviation change, i.e., 10 percentiles change in skill distribution. The instrument used for the ICT skill index, precited tower density, is derived from a district-level OLS regression that regresses district tower density on tower density in telecom circle leaving the district out while allowing for the relationship to vary across telecom circles and 2011 Census district characteristics. All the IV regressions control for the 2011 Census district characteristics that are used for the instrument construction. All models control for state-region fixed effects, individual and district-level controls reported in Table 1. Standard errors clustered at district-level are reported in parentheses. In different panels, interactions of ICT skills index are instrumented by interactions with predicted power density.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8: Mechanisms: Education and Job Search

	(1)	(2)	(3)	(4)	(5)	(6)
	Labor Force Participation (1/0)	Employed (1/0)	Regular salaried (1/0)	Self-employed (1/0)	Daily Wage work (1/0)	log (per capita expenditure)
Panel A: The Role of Education: Omit individual education dummies from the specification used in Table 4						
ICT skills index	-0.006	0.006	0.063***	-0.041*	-0.023	0.084*
	(0.015)	(0.015)	(0.018)	(0.021)	(0.014)	(0.049)
Observations	622,764	622,764	622,764	622,764	622,764	622,764
Kleibergen-Paap rk Wald F	8.524	8.524	8.524	8.524	8.524	8.524
Panel B: Testing Internet-based Job Search in Salaried Employment exploiting whether father is in salaried employment						
ICT Skills Index			0.045			
			(0.028)			
Father not in regular salaried *ICT Skills Index			0.023**			
			(0.010)			
Observations			126,227			
First Stage F Statistic:						
Sanderson and Windmeijer F, 1			9.174			
Sanderson and Windmeijer F, 2			44.11			

Notes: (1) The table reports the coefficient on ICT skill index from separate instrument variable (IV) regressions with different outcomes. A one unit change in ICT skill index means 0.254 Standard Deviation change, i.e., 10 percentiles change in skill distribution. (2) The instrument used for the ICT skill index, predicted cell tower density, is derived from a district-level OLS regression that regresses district tower density on tower density in the telecom circle leaving the district out while allowing for the relationship to vary across telecom circles and 2011 Census district characteristics. (3) All the IV regressions control for the 2011 Census district characteristics that are used for the instrument construction. All models control for state-region fixed effects, individual and district-level controls reported in Table 1. (4) In Panel B, interaction of ICT skills index is instrumented with interaction of predicted cell tower density. (5) Panel B restricts the sample to individuals whose fathers are in age 21-60 and coresident in the households, hence, Panel B refer to son/daughter-father sample. (6) Standard errors clustered at district-level are reported in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Online Appendix

Lightning and Cell Towers: An instrumental Variable for ICT skills in India?

Chiplunkar and Goldberg (2022) use lightning information as an instrument variable for 3G coverage. They construct region-level data (two or more time periods) from cross sectional surveys from 14 countries spanning across Latin America, Asia, and Africa. They calculate the average number of lightning strikes in a region between 2000-2014 and create a binary variable that takes the value 1 if the region receives above-median lightning strikes per square kilometer on average, and 0 otherwise. Controlling for region and year fixed effects in addition to other region-level controls, they show that regions that were categorized in above median lightning category experienced slower growth in 3G expansion over time establishing the first stage relation. They construct the lightning measure at region-level in the following ways:

$$Lightning_r = \sum_{p \in r} flashes_{pr}\, \theta_{pr,2001}$$

where $flashes_{pr}$ is measure of lightning in pixel p in a region r, while $\theta_{pr,2001}$ is fraction of individuals of that region who live in pixel p in 2001 ($\sum_p \theta_p = 1$). They use population data that is defined at the 1×1 km grid. In the appendix of their paper, they provide more details of lightning data. They use LIS/OTD 2.5 Degree Low Resolution Monthly Climatology Time Series (LRMTS) (Cecil, Buechler and Blakeslee, 2014; Cecil, 2006). Chiplunkar and Goldberg (2022, p47) state "Data is obtained as a raster with each grid corresponding to 2.5×2.5 degrees for each month, which is then averaged across months in a year." A 2.5° by 2.5° grid square at the equator covers an area of approximately 77,450 square kilometers, hence the lightning measure value is same for a very large area.

Given that the area of an average district in India is only about 5,000 square kilometers, the lightning measure from LRMTS is not very useful for us as it does not provide much variation, especially after controlling for state-region fixed effects (state-regions consists about 7-8 districts in a state). We use LIS 0.1 Degree Very High-Resolution Gridded Lightning Annual Climatology (VHRAC)[1] and LIS 0.1 Degree Very High-Resolution Gridded Lightning Full Climatology

[1] http://dx.doi.org/10.5067/LIS/LIS/DATA304

(VHRFC)[2] to construct two alternative measures of Lightning (Albrecht, Goodman, Buechler, Blakeslee, and Christian, 2016). A 0.1° by 0.1° grid square covers an area of 113.4 square kms at geographical center point in India (23.5120° N, 80.3290° E). Since an average district area is about 500 square kms, our lightning measure should vary not only across districts but also within districts for majority of districts (only 10 out of 640 2011 Census districts have area less than 114 square kms). Table A1 provides a description of climatology data discussed here.

Table A1. Climatology gridded dataset

File identifier	Dataset name	Description	Units	Dimensions	Bin sizes	Smoothing
VHRFC	VHRFC_LIS_FRD	Mean annual flash rate density	fl km–2 yr–1	760 × 3,600	0.1° × 0.1°	None
VHRAC	VHRAC_LIS_FRD	Mean daily flash rate density	fl km–2 day–1	760 × 3,600 × 365	0.1° × 0.1° × 1 day	49-day and 1° × 1° average
LRMTS	LRMTS_COM_FR	Monthly time series of flash rate (weighted)	fl km–2 day–1	72×144×240	2.5°×2.5°×1 month	Weighted average of LRTS_OTD_FR and LRTS_LIS_FR average

Source: Albrecht et al. (2016) and Cecil et al. (2014). In the dimensions, the first number represents latitude points, second represents longitude points, while the third represent time dimensions. 72×144×240 (for LRMTS) means that the flash values are reported for 10368 points (72×144 (latitude, longitude)) for 240 months.

VHRAC provides daily flash rate density for 365 days averaged over 1998 to 2013. We take an average over 365 days to get a measure of flashes per square kms per day. VHRFC provided flashes per square kms per year averaged over 1998-2013. No temporal or spatial smoothing is applied for VHRFC measure (Table A1). We construct our lightning measure in a similar way as done in Chiplunkar and Goldberg (2022).

$$Lightning_d = \sum_{p \in d} flashes_{pd}\, \theta_{pd,2020}$$

where $flashes_{pd}$ is measure of lightning in pixel p in a district d, while $\theta_{pd,2020}$ is fraction of individuals of that district who live in pixel p in 2001 ($\sum_p \theta_p = 1$). Our population data is 1×1 km grid downloaded from WorldPop. It is worth pointing out that $flashes_{pd}$ does not vary across each pixel (1×1 km grid) but could take few distinct values depending on the size of the district.

[2] http://dx.doi.org/10.5067/LIS/LIS/DATA301

Like Chiplunkar and Goldberg (2022), we also create indicators for above median districts using both of our flash measures. Table A2 provides the descriptive statistics of our lightning measures.

Table A2: Descriptive statistics of lightning measures

Number of districts	640
Lightning based on flashes per sq km per day	0.043 (0.025)
Lightning based on flashes per sq km per year	13.355 (9.054)

Note: Standard deviations are in parenthesis.

In Table A3, we provide the first stage results to check potential use of lightning measures as instrument in our context. In column (1) of Table A3, we regress ICT skills index on the first flash measure, flashes per sq km per day. The model also controls for state-region fixed effects, district-level, and individual-level controls described in Table 2 of main text. We do not find a significant correlation between ICT skills and lightning measure. Adding controls for average rainfall in district and district elevation does not make any difference (results not reported here). In column (2) of Table A3, we present the correlation between the ICT skills index and whether the district received above median flash controlling for same set of variables as column (1). We do not find any relation between ICT skills index and whether the district receives above median lightning. In column (3) and (4), we estimate similar relation as column (1) and (2), respectively, however, use the second measure of lightning, flashes per sq km per year. Results are similar and we fail to reject null correlation between ICT skills and lightning.

Table A3: First stage results with different Lightning measures as instruments

Dependent variable: ICT skills index	(1)	(2)	(3)	(4)
Lightning based on flashes per sq km per day	-1.787			
	(3.011)			
Indicator for above median districts, based on flashes per sq km per day		-0.075		
		(0.091)		
Lightning based on flashes per sq km per year			0.004	
			(0.007)	
Indicator for above median districts, based on flashes per sq km per year				-0.075
				(0.091)
Observations	622,764	622,764	622,764	622,764
R-squared	0.506	0.506	0.506	0.506

Note: Standard errors clustered at district-level are reported in parentheses. All models control for state-region fixed effects, individual and district characteristics as described in main text Table 2.

In Table A4, we explore whether our district level tower density correlates with the Lightning measures. Column (1) of Table A4 regresses district tower density on Lightning measure, yearly flashes per sq km.[3] As tower density may be potentially serially correlated within telecom circle, we cluster the standard errors at telecom circle level. We also report the robust standard error in the brackets to ensure clustering is not driving our main conclusion. The estimate in column (1) of Table A4 suggests some negative unconditional correlation between lightning and tower density. In column (2), we add the average rainfall and mean elevation of the district as controls.[4] Overall, the negative correlation remains. In column (3), we control district population density and share of district population living in urban areas. The column (3) correlation estimate is not only statistically insignificant, but magnitude of the conditional correlation shrinks towards zero. Hence, one can conclude that controlling for district population characteristics, there is no effect of lightning on district tower density. In columns (4) to (6), we explore the correlation between district witnessing above median Lightning and tower density. As evident from the estimate from column (6), the magnitude of correlation between above median lightning strikes

[3] We get similar conclusions with flashes per sq km per day as lightening measure.

[4] The rainfall data is extracted from NASA's IMERG Grand Average Climatology image that provides the average amount of precipitation that falls each year (mm/year), computed from June 2000 - May 2023 (https://gpm.nasa.gov/data/imerg/precipitation-climatology#grandaverageprecipitationclimatology). The elevation data is from development data lab's shrug database (https://www.devdatalab.org/shrug).

and tower density shrinks considerably besides being statistically insignificant once district population characteristics are controlled for.

Table 4: Correlation between tower density and Lightning measures

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: district tower density						
Flashes per sq km per year	-0.413	-0.649*	0.017			
	(0.255)	(0.329)	(0.157)			
	[0.206]	[0.246]	[0.106]			
Above median flashes				-6.246	-7.104	-0.568
				(3.706)	(4.140)	(2.224)
				[3.299]	[3.262]	[1.692]
Mean Elevation		0.009			0.008	
		(0.008)			(0.008)	
		[0.006]			[0.006]	
Average yearly rainfall		0.007			0.005	
		(0.006)			(0.006)	
		[0.003]			[0.003]	
District population density			0.000			0.000
			(0.001)			(0.001)
			[0.001]			[0.001]
Urban share in district population			84.376***			84.097***
			(13.282)			(13.280)
			[11.351]			[11.342]
Constant	24.646***	15.978**	-7.475**	22.480***	14.497*	-6.906***
	(5.418)	(7.089)	(2.643)	(4.074)	(7.615)	(2.245)
	[3.909]	[3.114]	[2.076]	[2.952]	[3.228]	[2.001]
Observations	640	640	640	640	640	640
R-squared	0.012	0.038	0.577	0.013	0.030	0.577

Note: The dependent variable for all models is district-level tower density. Standard errors clustered at the telecom circle level are reported in parenthesis, while robust standard errors are reported in brackets. All regressions are weighted by district population.

Chiplunkar and Goldberg (2022) argue that the equipment needed for mobile phone infrastructure is particularly sensitive to electrical surges caused by frequent lightning strikes, thus increasing expected costs and slowing the expansion of 3G internet. It should be noted that our endogenous variable differs from Chiplunkar and Goldberg (2022) and we exploit cross-section variation compared to theirs across time. Moreover, as discussed in the main text, the telecom sector in India has been highly competitive industry where multiple telecom operators are trying to garner a larger share of the market by providing access. As evident from the expansion of Base

Transceiver Stations (BTS) in recent years in India, perhaps the potential losses due to lightning strikes are probably less of a concern given the competition in the market. Moreover, the telecom companies can protect harm to BTS from lightning strikes by using methods such as lightning rods, grounding systems, surge protectors, etc.

REFERENCES

Albrecht, R., S. Goodman, D. Buechler, R. Blakeslee, and H. Christian. (2016). LIS 0.1 Degree Very High Resolution Gridded Lightning Climatology Data Collection. Data sets available online [https://ghrc.nsstc.nasa.gov/pub/lis/climatology/LIS/] from the NASA Global Hydrology Resource Center DAAC, Huntsville, Alabama, U.S.A. doi: http://dx.doi.org/10.5067/LIS/LIS/DATA306

Cecil, D. J. (2006). LIS/OTD 2.5 Degree Low Resolution Monthly Climatology Time Series (LRMTS). Dataset available online from the NASA Global Hydrometeorology Resource Center DAAC, Huntsville, Alabama, U.S.A. https:// doi. org/ 10. 5067/ LIS/ LISOTD/
DATA3 09

Cecil, D. J., Buechler, D. E., & Blakeslee, R. J. (2014). Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. Atmospheric Research, 135, 404-414.
Chiplunkar, G., & Goldberg, P. K. (2021). Aggregate implications of barriers to female entrepreneurship (No. w28486). National Bureau of Economic Research.

Mondal, U., Panda, S. K., Das, S., & Sharma, D. (2022). Spatio-temporal variability of lightning climatology and its association with thunderstorm indices over India. Theoretical and Applied Climatology, 149(1), 273-289.