# Who benefits from better Internet connectivity? Evidence from the labor market in South Africa

Zhe Liu *

September, 2022

**Abstract**

I study how job outcomes and search channels used in the labor market respond to the expansion of Internet availability. I use two-way fixed effects identification strategy with continuous treatment at district level, and find that Internet availability has a positive impact on average employment and total income. After Internet access improves in their areas, job seekers are more inclined to check for job information online than using personal networks. Unskilled workers are discouraged from searching online, less likely to be employed, and earn less. Young workers search through more methods and increasingly rely on personal networks, but are paid less than the experienced. Constraints on effective uses of Internet job search, and other Internet activities such as social networking could help explain the results.

**Keywords:** Unemployment, Job Search, Internet, Social Networks

*Zhe Liu is a PhD candidate in the Economics department at University of Washington (zliu525@uw.edu). This is a preliminary draft. Please do not cite or distribute without author's permission.

# 1 Introduction

A lack of information is arguably one of the key frictions in labor markets, and growing evidence has shown that information frictions can impede transitions into employment (Caria, Lessing and Hermes, 2019). In many developing and emerging economies, individuals often rely on informal sources, such as family and neighbors, or are left with no information when making labor market decisions. Referrals or reference letters are used as important methods for filling vacancies and finding jobs in these countries (Abel, Burger and Piraino, 2020; Beaman and Magruder, 2012). In recent years, the growth of Internet adoptions and expansion of job sites have lowered the cost of acquiring and disseminating job related information (Autor, 2001; Kuhn and Skuterud, 2004). Internet-based job search is by now one of the predominant ways of searching for jobs (Kuhn and Mansour, 2014). In a standard labor market searching model, if individuals have cheap access to more information about vacancies, they will be able to consider more options and thus find a better job or be more likely to find a job. Productivity and wages should rise. This additional low-cost information should better prepare actors to transact in the market closer to perfect information.

An interesting question is to what extent can a market mechanism like online job search and hiring, open to all and anonymous, substitute for exclusionary personal connections. A consensus estimate that at least half of all jobs are typically found through informal contacts (e.g. personal networks) rather than through formal search methods (Topa, 2011). A rich and growing literature has been documenting the use of personal networks, its impacts on labor market outcomes, and usage variation across demographic groups (Arbex, O'Dea and Wiczer, 2019; Beaman, Keleher and Magruder, 2018; Calvó-Armengol and Jackson, 2004; Granovetter, 1973; Heath, 2018; Holzer, 1988; Montgomery, 1991; Mouw, 2003; Pellizzari, 2010).

Answering this question in South Africa has important policy implications. Being Africa's most industrialized economy, South Africa still has extremely high levels of unemployment. The recent Covid-19 lockdown has pushed the unemployment rate to a record high above 30%[1]. In particular, unemployment has been inordinately high for young workers (Figure A1), who have less access to referral networks and limited information about their employment prospects.

Existing studies solely focus on Internet impacts on job outcomes such as employment rate and income. My study aims to contribute by providing evidence on job search efforts and information channels. In this paper, I estimate how broadband Internet availability affects job outcomes for workers with different skill levels and age groups in South Africa, and how search methods used by workers and firms respond to faster and easier Internet access. To do so, I use two-way fixed effects (TWFE) and compare individuals in locations with different Internet penetration rates, before and after the arrival of first undersea cables in South Africa in 2009. To the best of my knowledge, this is the first paper examining whether Internet access has an impact on the choice of job search methods.

I present a simple model of job seekers' utility maximization to show that search effort is the key to how employment changes with Internet access. Comparative statics predicts that if more Internet availability brings down the cost of search and increases the marginal productivity of search, a job seeker will search more and has higher probability of being employed. If firms have already adopted a high amount Internet in the hiring process, more searches may lead to excess applications that firms cannot match with vacancies in time, the marginal productivity of search in response to more Internet can be negative in this case. Then the impact on optimal search effort is ambiguous, and will depend on the balance between the marginal productivity of search and the cost of search.

I match Internet connection data with spatially coded panel data of job search activities from South Africa, and find positive effects of Internet availability on employment and total income. A one-standard-deviation increase in Internet availability (about 10 percentage points) increases the employment rate for an average jobseeker in the district by 3.6 percent, and increases his or her total income by 3.5 percent. More Internet availability induces jobseekers to use online job information by about 10 percent, while decreasing their reliance on personal networks by less than 1 per-

[1] Source: Statistics South Africa

cent. The total number of different search methods used decline, suggesting lower search effort is made by the job seekers with more Internet access. I show that the estimates do not change appreciably if I include a set of time-varying controls for potential productivity factors, and if I allow for different time trends across areas.

Heterogeneous analysis by age group and education attainment contribute to our understanding of distributional effects of Internet technology change. Compared with experienced workers when more Internet becomes available in the area, young workers (between 15 to 24 years old) share similar probability of being employed but with large income cuts. Young workers will use more different search channels, and more likely to search through their personal networks. Compared with skilled workers with primary or more education, unskilled workers are discouraged from online job search, less likely to be employed, and earn less.

I provide evidence on possible mechanisms on how internet access may affect employment and search channels across age groups and skill levels. Constraints on effective online search like computer literacy, and various Internet activities like social networking could help explain some of the findings on choices of search methods. The results on employment and income should reflect the equilibrium outcomes of both labor supply and labor demand. Without employers or firms' data, I cannot say much about how employers' job creation decisions respond to more Internet access empirically. Instead, I use the standard search and matching theory of unemployment and vacancies by Mortensen (1986), and simulated the expected general equilibrium impacts. The results show that Internet can change the equilibrium outcomes by decreasing the job search costs for workers, or increasing matching efficiency for firms.

To help assess the causal effect of Internet access, I address the endogeneity issues in two ways. First, I analyze the timing of changes in Internet and employment, and find that the timing does not appear to be systematically related to key observable correlated of employment. Past levels of productivity variables does not predict current Internet connection levels. Second, I use both location fixed effects and year fixed effects to explore the temporal and spatial variation in the Internet availability across 52 districts in South Africa. This identification is similar in spirit to a difference-in-difference (DID) design at district level with a continuous treatment. Variations in Internet treatment intensity make it possible to evaluate a "does-response" relationship, and can bolster the results for a causal interpretation. In addition, policy makers may care more about the effect of changes in availability rates than about the effect of the existence of Internet. Recent literature shows that TWFE estimators are not robust to heterogeneous treatment across groups and over time (Goodman-Bacon, 2021), especially with a continuous treatment design. So I also assess the robustness of the TWFE estimators following de Chaisemartin and D'Haultfœuille (2020).

This paper adds developing country evidence to a limited literature assessing the linkages between Internet and labor market outcomes. More than 60 per cent of the world's employed population earn their livelihoods in the informal economy, most of them in emerging and developing countries (Bonnet, Vanek and Chen, 2019). It remains questionable if findings for broadband Internet expansion implemented in developed countries are applicable to developing countries. To date, the only direct evidence on the average and distributional effects in developing countries is provided by Hjort and Poulsen (2019) focusing on the Africa continent. They leverage the gradual arrival of sub-marine Internet cables in Africa, and find large positive effects on employment and incomes, particularly for higher-skill occupations. This happens in part due to the technology's impact on firm entry, productivity, and export. Existing studies in high-income countries show mixed impacts. Kroft and Pope (2014) analyze the expansion of Craigslist in the US, and find that Craigslist significantly lowered classified job advertisements in newspapers, but had no effect on the unemployment rate. Dettling (2017) uses state-wide shares of multifamily residences to instrument for the diffusion of Internet access across the U.S., and finds increases in labor force participation rates of married women, and no corresponding effect for single women or men. Bhuller, Kostol and Vigtel (2019) document that broadband expansions in Norway increase online vacancy-postings, lower the average duration of a vacancy, resulting in higher job-finding rates and starting wages, and more stable employment relationships after an unemployment-spell. Akerman, Gaarder and Mogstad (2015) finds the same Internet expansion improves the labor market outcomes and productivity of skilled workers only.

My paper complements a growing experimental literature considering the role of limited information in labor market matches in developing countries. While my paper evaluates the general equilibrium impacts of cheaper information, field experiments have the advantage of being able to isolate and disentangle information interventions targeted at either jobseekers or firms or both. Abebe et al. (2021) shows that job application workshop for young jobseekers can help them signal skills better, and generate large and persistent improvements in their labor market outcomes. The effects are larger when combined with formal certificates provided to firms (Carranza, Garlick and Orkin, 2020). Firms may have poor knowledge of candidates, and providing information directly to firms can improve match quality (Abel, Burger and Piraino, 2020; Banerjee and Chiplunkar, 2020). Online platforms such as LinkedIn can help address supply-side information frictions by allowing jobseekers to learn more about job prospects, and also address demand-side frictions by allowing firms to learn more about potential candidates (Wheeler et al., 2022). My findings on the distributional effects are at odds with the notion that active labor market programs such as training or employment subsidies have larger employment effects for more disadvantage groups(Card, Kluve and Weber, 2018).

This paper also extends the literature on the role of information and communications technology (ICT) in developing countries. ICT such as mobile phones has been attributed with reducing price dispersion across markets and increasing welfare for producers and consumers (Aker and Mbiti, 2010; Goyal, 2010; Jensen, 2007). ICT such as mobile money can help reduce transaction costs and potentially improve informal risk sharing networks (Jack and Suri, 2014). ICT can even influence fertility patterns and bring cultural changes to the society (La Ferrara, Chong and Duryea, 2012). My paper shows that ICT such as Internet can provide cheaper access to job information, or reduce communication costs with family and friends, and improve employment outcomes for job seekers in the labor market.

The rest of the paper proceeds as follows. Section 2 provides a conceptual model analyzing how Internet availability affects job seekers' search effort and employment. In Section 3 I present the data, and in Section 4 the two-way fixed effects estimation strategy. The average and heterogeneous results are in Section 5. In Section 6, I explore mechanisms how Internet access may affect employment and search behavior. Section 7 concludes with policy implications.

# 2 Conceptual model

I present a simple model illustrating the relationship between employment, job search, and Internet access. A job seeker lives two periods: in the first period, an unemployed individual receives some unemployment benefit $b$ and decide how much effort to spend for job searching $s$. Cost of job search $\tau(\theta)$ depends on amount of Internet access, and the probability of finding a job depends on both the search effort and amount of Internet access: $p(s,\theta)$. In the second period, if the individual becomes employed, assuming labor supply is inelastic, a fixed income will be given as $w$. In this set up, internet access amount $\theta$ and wage $w$ are given exogeneously. The job seeker chooses job search effort and maximizes the expected lifetime utility as follows:

$$\begin{aligned} \max_{s} \quad & u\left(c_1\right)+\beta E u\left(c_2\right) \\ \text{s.t.} \quad & c_1=b-\tau(\theta) s \\ & c_2=\begin{cases} w & \text{w.p. } p(s,\theta) \\ b & \text{w.p. } 1-p(s,\theta) \end{cases} \\ & 0 \leq p(s,\theta) \leq 1 \end{aligned} \tag{1}$$

where utility $u(c)$ is assumed to be increasing and strictly concave in consumption. An interior solution should satisfy the following first order condition:

$$\tau(\theta) u^{\prime}=\beta \frac{\partial p(s,\theta)}{\partial s}\left[u\left(w\right)-u\left(b\right)\right] \tag{2}$$

which implies that the individual chooses search effort $s$ optimally such that the marginal utility of giving up consumption equals the expected utility gain from searching for work, which is the difference between employment and unemployment utility in the second period.

For this paper, I am interested in how employment probability may change with the Internet access given exogeneously. The comparative statics is,

$$\frac{d}{d\theta}p(s(\theta),\theta) = \frac{\partial p}{\partial s}s'(\theta) + \frac{\partial p}{\partial \theta} \tag{3}$$

Assuming the marginal productivity of search and Internet are both positive ($\frac{\partial p}{\partial s}, \frac{\partial p}{\partial \theta} > 0$), the effect on employment will depend on $s'(\theta)$. In order to see how optimal search effort $s^{\star}(\theta)$ changes with Internet access $\theta$, we can differentiate the first order condition equation 2 with respect to $\theta$:

$$s'(\theta) = \frac{\tau' u' - \beta p_{s\theta}(u^{emp} - u^{unemp})}{\tau u'' + \beta p_{ss}(u^{emp} - u^{unemp})} \tag{4}$$

where $u^{emp}, u^{unemp}$ represent the utility being employed and unemployed in period 2 respectively.

Assume $u'' < 0$ and $p_{ss} < 0$, and u$^{emp} > u^{unemp}$ is a necessary condition for the existence of an interior solution, the denominator in equation 4 is negative. The sign of the numerator depends on two parts. First, $\tau'(\theta)$, the change in the cost of job search given more Internet access. If we think more Internet means that jobseekers have cheaper access to more job related information, the cost of job search should be lower, $\tau'(\theta) < 0$. Second, $p_{s\theta}$, the change in the marginal productivity of search in response to more Internet access. $p_{s\theta} > 0$ if job search by the jobseekers is made more productive with more Internet, eg. Internet technology can help firms screen candidates and match them with vacancies faster. Then the optimal search effort $s^{\star}(\theta)$ should be positive, and the change in employment probability should be positive too. However, the marginal productivity of search $p_s$ is not necessarily linear in $\theta$. For example, if firms have adopted a high amount of Internet $\theta$ in their hiring process, more searches will create excess applications that firms cannot process in time. Thus, the marginal productivity of search in response to more Internet access can be negative, $p_{s\theta} < 0$. Then the impact on optimal search effort is unclear.

This conceptual model shows how employment changes with Internet access depends on optimal search effort. If search effort increases with Internet access

($s'(\theta) > 0$), equation A3 indicates that employment will increase as well; however, if search effort declines with Internet access ($s'(\theta) < 0$), the net impact on employment is unclear. Theoretically, if search cost decreases while marginal productivity of search increases with more Internet, then job seekers will exert more effort. However, it is also possible that marginal productivity of search declines, then how search effort changes is ambiguous and depends on the balance between the factors mentioned above. Empirically, I test the change using the total number of different search methods as a proxy measurement for search effort in section 5.

Internet can also change the utility of leisure, which will impact the trade off between searching for jobs and staying unemployed. I solve a version of this model including leisure in the job seeker's utility function in appendix A. The comparative statics predictions are similar.

Using published data from a field experiment that Abel et al. (2019) have done with South Africa youth, I also find suggestive evidence that online job search is correlated with higher effort exerted. The original experiment is to test the effects of plan making on job search and employment. Table 1 shows the regression results using panel data over two follow-up periods, with only baseline control group observations included. In a period of about 12 weeks, individuals who search jobs online spend 2 hours more, and send out 2.6 more number of applications in total. They are also more likely to receive responses and job offers from the employers.

Table 1: Effects of Online Search on Search Behaviors and Employment Outcomes

| | (1)<br>Search Hours | (2)<br>Applications | (3)<br>Empl Responses | (4)<br>Job Offers | (5)<br>Employed |
|---|---|---|---|---|---|
| Search online | 2.091 | 2.626*** | 0.535*** | 0.091*** | 0.016 |
| | (1.300) | (0.320) | (0.061) | (0.026) | (0.031) |
| Mean Dep Vars | 14.087 | 3.821 | 0.543 | 0.131 | 0.116 |
| Obs | 818 | 828 | 828 | 819 | 857 |
| R-squared | 0.026 | 0.079 | 0.048 | 0.026 | 0.011 |

* Notes: All specifications control for age, gender, education, and location fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# 3 Data

In the main analysis, I use the South Africa National Income Dynamic Studies (NIDS) for labor market data. NIDS is the first and only national household panel survey in South Africa, and is implemented by the Southern Africa Labour and Development Research Unit (SALDRU) based at the University of Cape Town's School of Economics. The study began in 2008 with a nationally representative sample of over 28,000 individuals in 7,300 households across the country. The core survey continued to be repeated with these same household members every two years to three years, with the latest interview round being conducted in 2017. NIDS provides information about changes in these broad themes, including poverty, education, health, household structure, labor market participation and economic activity, migration, and social capital.

I focus on the labor market module in the survey, where working age adults were asked about their labor market participation and economic activity, including employment status, income (wages or the profits of self-employed workers), contract types, and industry. In addition, individuals were asked to check all the job search methods used, including family and friends, online ads, government agency, previous employers, and others.

Table 2 provides the summary statistics of working age (15 - 65) individuals used in the analysis. This sample has a balanced representation of urban and rural population. 59 percent of the sample are female. Average worker's age is 33, and 37 percent are considered as youth workers. 52 percent have finished primary education. Cellphone ownership is high (71 percent) compared to computer ownership (5 percent). About one third of the sample report they know how to use a computer. 37 percent of the sample are employed, among which 27 percent have a job paid with regular salary, and 5 percent are self employed. On average, individuals work around 40 hours per week, and earn 3233 ZAR(230 USD) per month. The standard deviation of log income is large, because I put zero for unemployed individuals' income. Network (25 percent) is the most widely used job searching method, while 6 percent of the sample report that they have used online search. Internet is available to 10 percent of the population in an average district with a standard deviation of 14 percentage points.

I use the Internet infrastructure and speed data published in Hjort and Poulsen (2019). Using Mahlknecht's map of submarine cables to measure landing points and times (Mahlknecht 2014), and www.africabandwidthmaps.com and AfTerFibre's (AfTerFibre 2014) maps of terrestrial backbone networks to measure locations' connectivity, the authors document whether a city is connected to the Internet quarterly from 2007 to 2014. Average Internet speed for the same locations is also provided by network service company Akamai Technology.

I match the Internet connection data from Hjort and Poulsen (2019) with the NIDS survey data using the geocode and year. While the converge data are available at the city level, individuals in the NIDS outcomes data can only be identified at higher levels of geographies, such as province and district. I aggregate the city-level connection data to district-level by calculating the percentage of cities with connection in one district by year, weighted by its population. 52 districts across 4 waves from 2008 to 2014 are matched.

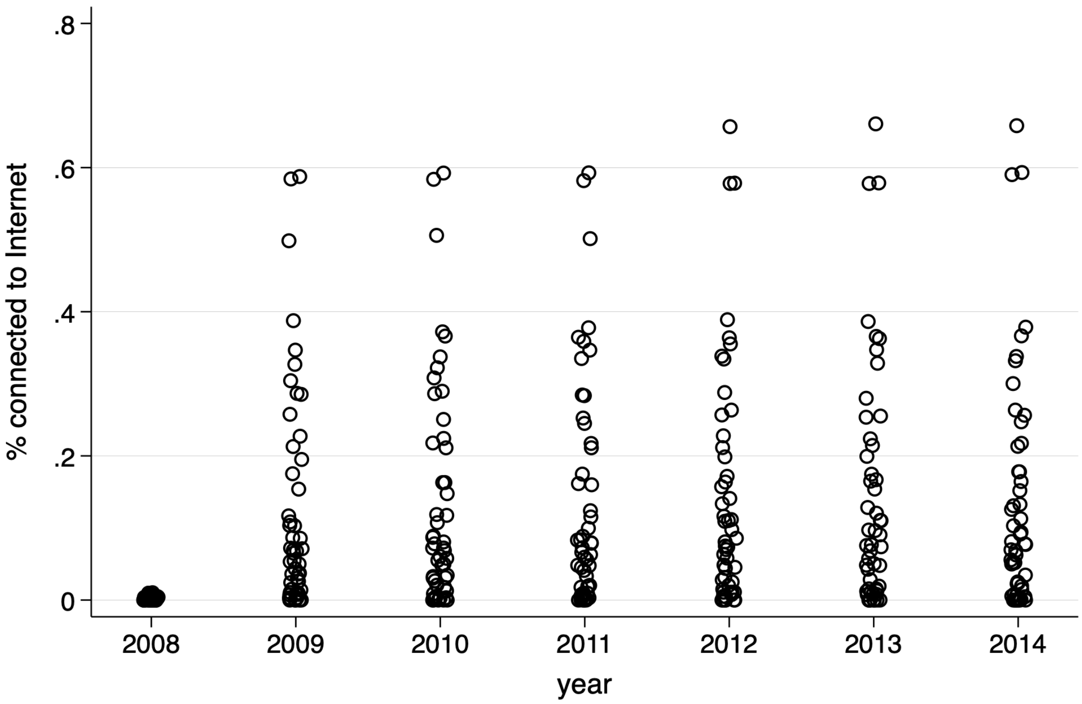
Figure 1 and 2 show the variation in percent of cities connected over time and across districts. In 2008, all cities have no connection. Over the years, more cities gained access and more districts achieved higher availability rates in 2014. There are also differences in connection timing and access intensity within districts, which generate a continuous measure of availability rates that I exploit as the key variations in my empirical analysis.

Table 2: Sample Summary Statistics

| | Obs | Mean | SD |
|---|---|---|---|
| *Individual characteristics* | | | |
| Urban area | 38,497 | 0.47 | 0.50 |
| Age | 38,520 | 33.49 | 14.29 |
| Female | 38,520 | 0.59 | 0.49 |
| Youth(15-24) | 38,520 | 0.37 | 0.48 |
| No school | 38,436 | 0.06 | 0.23 |
| Primary education | 38,436 | 0.39 | 0.49 |
| Secondary education | 38,436 | 0.29 | 0.45 |
| Tertiary education | 38,436 | 0.26 | 0.44 |
| Parents with primary education | 24,156 | 0.19 | 0.39 |
| Own a cellphone | 35,311 | 0.71 | 0.46 |
| Own a computer | 35,301 | 0.05 | 0.22 |
| Is computer literate | 34,386 | 0.29 | 0.46 |
| *Household characteristics* | | | |
| HH owns a cellphone | 38,124 | 0.85 | 0.36 |
| Spent money on cellphone monthly | 29,108 | 0.74 | 0.44 |
| HH owns a computer | 38,069 | 0.11 | 0.31 |
| Spent money on internet monthly | 29,144 | 0.01 | 0.11 |
| *Labor market outcomes* | | | |
| Employed | 37,108 | 0.37 | 0.48 |
| Salary job | 35,654 | 0.27 | 0.44 |
| Self employed | 35,651 | 0.05 | 0.21 |
| Total income (adjusted) | 32,475 | 2.34 | 3.85 |
| Salary income(adjusted) | 34,696 | 3.69 | 6.57 |
| Has permanent duration | 9,307 | 0.54 | 0.50 |
| Weekly hours | 10,516 | 39.70 | 17.15 |
| *Job search methods* | | | |
| Network | 32,923 | 0.25 | 0.43 |
| Online | 32,923 | 0.06 | 0.24 |
| Government | 32,923 | 0.03 | 0.17 |
| Others | 32,923 | 0.15 | 0.36 |
| *Internet connection at district level* | | | |
| % population connected | 38,520 | 0.10 | 0.14 |

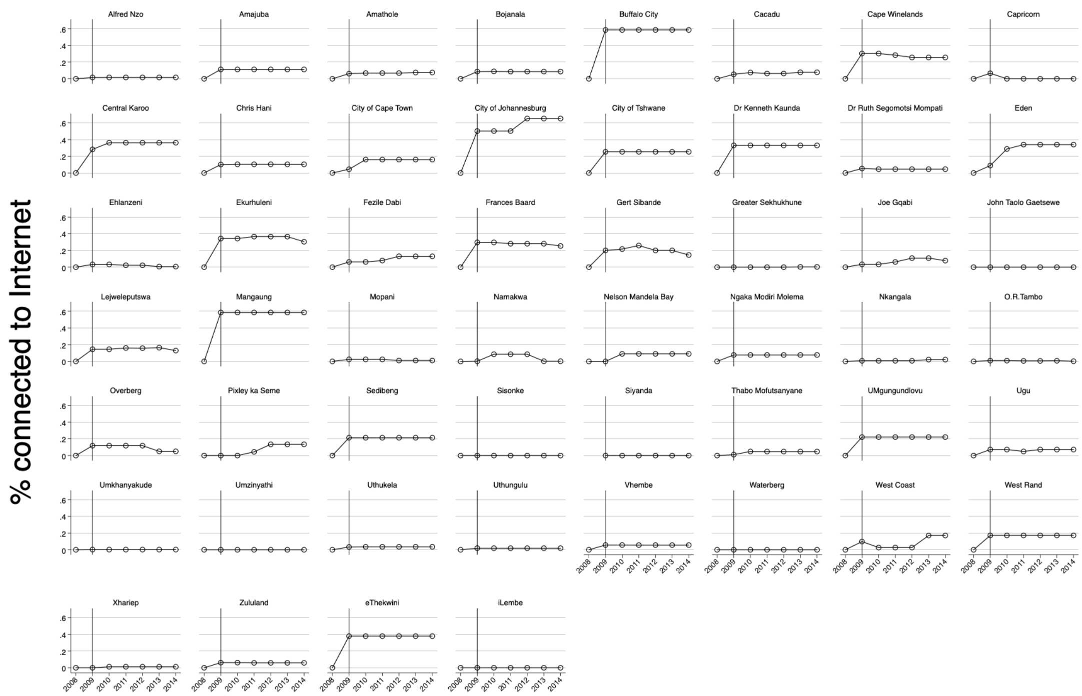
[*] Notes: Only workers between age 15 and 65 are included. The income for unemployed workers is adjusted as zero, and inverse hyperbolic sine is used for the log transformation.

Figure 1: Comparing Internet availability rates over years




Notes: Each dot represents one single district.In 2008, the percent of populations connected to Internet are zero for all districts. In 2009, some districts start to have higher connectivity. More districts have moved up and achieved higher availability rates by 2014.

Figure 2: Comparing Internet availability rates across districts




Notes: Each sub plot represents one district.The vertical line indicates the connection of first undersea cable in 2009.

# 4 Empirical Strategy

My empirical approach is a two-way fixed effects estimation that controls for location and year fixed effects. I compare individuals across locations with varying degrees of Internet coverage, before and after the connection of undersea Internet cable in 2009. This is motivated by two features of this Internet expansion. First, most of the confounding supply and demand factors are accounted for by the location fixed effects. Second, the timing of the expansion is unlikely to co-vary with key correlates of employment. In fact, I find 68 percent of the variation in Internet availability can be attributed to time-invariant location characteristics and common time effects, while 8 percent of the variation can be attributed to a set of time-varying variables.

### 4.1 Two-way Fixed Effects estimation

I run the following two-way fixed effects estimation as the main specification.

$$Y_{ijt} = \beta PercentConnected_{jt} + X'_{ijt}\alpha + \gamma_t + \theta_j + \epsilon_{ijt} \tag{5}$$

where $Y_{ijt}$ is the labor market outcomes for worker i in district j at time t. The set of outcomes of interest are individual-level labor market outcomes, including employment, employment with formal contracts, income, network search, and online search. $PercentConnected_{jt}$ is the percent of population in district j connected to the Internet at time t, or the Internet penetration rate. This measure allows me to exploit variation within the set of connected districts in their intensity of treatment.

All specifications include both location fixed effects, $\theta_j$, time fixed effects,$\gamma_t$, and an idiosyncratic error term, $\epsilon_{ijt}$. District fixed effect controls for any time-invariant differences in employment outcomes that may be correlated with access to the Internet, and year fixed effect controls for any within-district invariant changes in employment outcomes that may be endogenous to Internet access.

$X_{ijt}$ is a vector of individual- specific controls, including age,gender, and education level. Since there could be unobserved individual-level factors that are endogenous to the choice of search channels, I also include a individual fixed effect in some analyses. In all analyses, standard errors are clustered at the district level.

Within such a set up, as long as there are not omitted idiosyncratic shocks correlated with both Internet penetration rate and labor market outcomes, the causal effect of Internet, $\beta$, is identified off of comparison between the change in outcomes for locations that gain (more) access to Internet in a given year and the change in outcomes for other locations that without or gain less access at the same time.

Given that I am controlling for fixed effects for districts and years, the core of this design is similar to a difference-in-difference setup at the district level. Districts fixed effects act as controls for the "preperiod" outcomes of workers in the same area that never received Internet, which is the first difference. Treatment and control groups can be defined as workers within a year that had different exposure to Internet access. Subtracting the treatment group outcomes from the control group yields the second difference. Because the treatment variable $PercentConnected$ is continuous, I effectively weight these double differences by the difference in Internet penetration rates.

## 4.2 Tests for parallel trend assumption and timing of the expansion

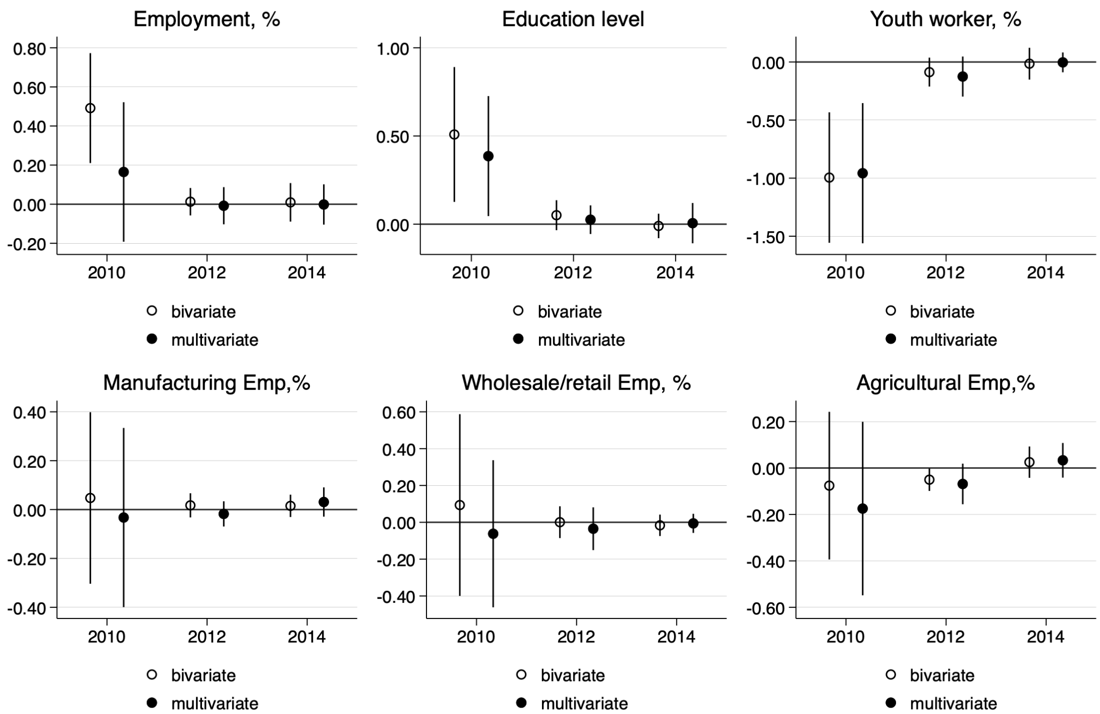
Another threat to identification is that the timing of the expansion might be related to different underlying trends across locations. To examine the timing of the expansion, I test if the Internet allocation is determined by baseline productivity variables, including employment rate, industry distribution, age, and education level. The estimating equation is as follows:

$$\Delta PercentConnected_{jt} = \gamma_t + \theta_t c_{j,t0} + \epsilon_{ijt} \tag{6}$$

where $\Delta PercentConnected_{jt}$ is the change in Internet availability, $\gamma_t$ is year fixed effect, and $c_{j,t0}$ is baseline district-level variables related to productivity, including employment rate, education level, percent of young workers, and industry distribution.

Figure 3 plots the estimated coefficient $\theta_t$ from equation 6, using both bilateral and multilateral model specifications. The expansion was larger in areas with higher employment rate, more educated and older population at the beginning when the whole country got connected to the fast Internet in 2009. From 2010 and onward, there appears to be little systematic relationship between the timing of the expansion and district characteristics.

Figure 3: Test for the Timing of Internet Expansion




# 5 Results

## 5.1 Main effects

In Table 3, I show the regression results for specification in 5, including district and year fixed effects and demographic controls. I find that one standard deviation increase in Internet connection (about 10 percentage point) increases the probability that an individual is employed by 1.3 percentage point, or 3.5 percent increase off a baseline of 36.6 percent average employment rate (column 1). This result is similar in magnitude to what Hjort and Poulsen (2019) find about South Africa in their cross-country sample of Africa countries. Total income will increase by 8 percent with additional Internet (column 2) [2].

To see what extent these increases reflect additional economic activity, I use more detailed work-related questions that only employed individuals were asked in the NIDS. Given the truncation by survey design, results in column 3-5 should not be interpreted as casual effects of Internet, but rather should be viewed as an intensive channel of the overall effects. For individuals already working, they will earn more while work less hours with additional Internet (column 3, 4). The estimated effect on having a formal contract is close to zero and insignificant (column 5). This helps rule out the situation that the additional employment comes from formalization of existing informal jobs.

To further explore how individuals job search behavior might change with Internet access, I show results on the search methods in Table 4. Additional 10 percentage point increase in Internet availability will induce jobseekers to look for information online by 0.64 percentage points, which is about 10 percent increase from the mean (column 1). Internet's negative impact on network search is small and not statistically significant, suggesting that network channel can be resilient to the Internet access shock (column 2). With more Internet access, dependencies on government agencies for job search increases by about 1.3 percentage point (column 3). The number of different search methods used by the individuals declined by 3.6 percent (column 4). If this total number of methods can be viewed as a proxy for search effort, this result could suggest that Internet access leads to lower search effort.

[2] 60 percent of the observations are reported not employed and not earning any income, and I put zero as their income.

Table 3: Fixed Effects Estimates of Internet Connection on Job Outcomes

| Outcome | Employed (0/1) (1) | Total income (asinh) (2) | Salary wage (asinh) (3) | Weekly hours (asinh) (4) | Formal contract (0/1) (5) |
|---|---|---|---|---|---|
| % connected | 0.135** | 0.836** | 1.610*** | -0.199 | 0.007 |
| | (0.065) | (0.343) | (0.521) | (0.162) | (0.078) |
| | [0.068] | [0.047] | [0.032] | [0.204] | [0.932] |
| Mean of outcome | 0.366 | 2.340 | 3.692 | 3.533 | 0.688 |
| Observations | 37,034 | 32,411 | 10,487 | 10,487 | 9,319 |
| R-squared | 0.160 | 0.190 | 0.162 | 0.076 | 0.121 |
| Year FE | Y | Y | Y | Y | Y |
| Location FE | Y | Y | Y | Y | Y |
| Controls | Y | Y | Y | Y | Y |

* Notes: Only workers between age 15 and 65 are included. Employed equals to 1 if the individual is employed with a salary job or self-employed. Hours and income are summed across each of the individual's jobs if more than one is reported. Total income are calculated using monthly income if salary employed, profit if self-employed, and as zero if unemployed. Inverse hyperbolic sign transformation are done to total income and salary wage. Only employed individuals are asked about wage, working hours, and contract types, so the number of observations for column 3-5 are small. Control variables include age, gender, and education level.Standard errors (in parentheses) are clustered at the district level. Given the small number of clusters, the wild bootstrap p-values [in brackets] are calculated following Cameron, Gelbach and Miller (2008). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Fixed Effects Estimates of Internet Connection on Search Methods

| Outcome | Online | Network | Government | Number of search methods |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| % connected | 0.064*** | -0.017 | 0.013 | -0.091* |
| | (0.017) | (0.064) | (0.016) | (0.051) |
| | [0.009] | [0.803] | [0.436] | [0.062] |
| Mean of outcome | 0.061 | 0.247 | 0.030 | 0.251 |
| Observations | 32,855 | 32,856 | 32,856 | 38,436 |
| R-squared | 0.085 | 0.053 | 0.022 | 0.031 |
| Year FE | Y | Y | Y | Y |
| Location FE | Y | Y | Y | Y |
| Controls | Y | Y | Y | Y |

* Notes: Only workers between age 15 and 65 are included. Network, Online and Government variables are equal to 1 if workers have used this method when searching for jobs. "Number of methods" is the total number of different methods used by the individual in the past four weeks. Control variables include age, gender and education level. Standard errors (in parentheses) are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## 5.2 Robustness checks

I show several specification checks to investigate the robustness of previous results in Table 5.

The first set of robustness checks examines whether the timing of the broadband Internet rollout correlates with time-varying covariates and/or trends. Column 1 shows the results without any controls, and column 2 repeats the main results with time-varying covariates. In column 3, I include linear trends interacted with baseline (year 2008) demographic covariates. In column 4, I allow for municipality-specific linear trends. Second set of checks include individual fixed effects since the choice of search channel could be endogenous (column 5). The point estimates are similar across these specifications.

Table 5: Main results robustness checks

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Employment | 0.109* | 0.135** | 0.102 | 0.175* | 0.101 |
| | (0.063) | (0.065) | (0.065) | (0.099) | (0.079) |
| Total Income | 0.649** | 0.836** | 0.609* | 0.933 | 0.481 |
| | (0.317) | (0.343) | (0.347) | (0.671) | (0.373) |
| No.of Methods | -0.085* | -0.091* | -0.140** | -0.023 | -0.221*** |
| | (0.049) | (0.051) | (0.054) | (0.145) | (0.074) |
| Online | 0.067*** | 0.064*** | 0.060*** | 0.079** | 0.027 |
| | (0.019) | (0.017) | (0.018) | (0.036) | (0.019) |
| Network | -0.029 | -0.017 | -0.063 | 0.093 | -0.055 |
| | (0.063) | (0.064) | (0.064) | (0.080) | (0.065) |
| Observations | 32,923 | 32,856 | 32,817 | 32,856 | 32,923 |
| Location FE | Y | Y | Y | Y | Y |
| Year FE | Y | Y | Y | Y | Y |
| Time-varying covariates | | Y | Y | Y | |
| Trends interacted with | | | | | |
| baseline covariates | | | Y | Y | |
| location FE | | | | Y | |
| Individual FE | | | | | Y |

* Notes: Each cell is from a separate regression. Standard errors (in parentheses) are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The second set of robustness checks are related to recent research on TWFE with heterogeneous treatment effects. TWFE regressions are unbiased for an ATE only if the treatment effect are constant between groups and over time. With heterogeneous treatment effects and under a parallel trends assumption, TWFE may estimate a weighted sum of treatment effects across periods and units, with some negative weights.The negative weights could bias the treatment coefficient in TWFE regressions close to zero or negative, even if the treatment effect is positive for every unit $\times$ period (de Chaisemartin and D'Haultfœuille, 2022; Goodman-Bacon, 2021). Several alternative difference-in-difference (DID) estimators robust to heterogeneous effects have been proposed recently. Most of them apply to binary treatments that follow a staggered design (Borusyak, Jaravel and Spiess, 2022; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). Only one estimator $DID_m$ proposed by de Chaisemartin and D'Haultfœuille (2020) can be extended for continuous treatments, which applies to my research design. The $DID_m$ estimator is a weighted average, across treatment intensity $d$ and period $t$, of DIDs comparing the $t-1$ to $t$ outcome evolution of groups whose treatment goes from $d$ to some other value, and of groups with a treatment equal to d at both dates, normalized by the intensity of the treatment change experienced by the switchers.

I first estimate the weights attached to TWFE estimator $\hat{\beta_{fe}}$, and find that 45 percent are positive, 55 percent are negative. The negative weights sum to -0.09. The correlation between the weights attached to $\hat{\beta_{fe}}$ and the year t is equal to 0.08 (t-stats = 9.7), suggesting that the effect of Internet may be different in the early years than in the later years of the panel. Given the possibility of heterogeneous continuous treatments, I compute the robust $DID_m$ estimator. Table 6 shows the $DID_m$ estimates share the same sign but with larger effects (col 1) from TWFE estimates as in 5. Less observations are used to avoid bad comparison between later treated and early treated groups (col 3). The placebo estimator $DID_m^{pl}$ compares the job outcomes and search channels of workers with and without an internet boost in their districts one period before the change, which can be used to test the common trend assumption. $DID_m^{pl}$ is large but insignificant, suggesting that workers in districts where the internet availability changed between t-1 and t may start experiencing a different pretrend from t-2 to t-1 than workers in districts where availability stays the same.

## 5.3 Heterogeneous effects by education level

Considering how Internet can be a skill-biased technology as documented in many rich countries (Akerman, Gaarder and Mogstad, 2015; Michaels, Natraj and Van Reenen, 2014), I test if this is true in South Africa by interacting Internet penetration rate with educational attainment as shown below. The education level is

Table 6: DID estimators robust to heterogeneous treatment effects

| | $DID_m$ | | | Placebo $DID_m^{pl}$ | | |
|---|---|---|---|---|---|---|
| | Estimates (1) | SE (2) | N (3) | Estimates (4) | SE (5) | N (6) |
| Employment | 0.170 | 0.255 | 20306 | 0.043 | 0.622 | 977 |
| Total Income | 0.692 | 1.500 | 15744 | -0.180 | 1.02 | 766 |
| No. of Methods | -0.111 | 0.365 | 21914 | -0.194 | 0.929 | 1128 |
| Online | 0.019 | 0.068 | 16349 | 0.078 | 0.106 | 785 |
| Network | 0.132 | 0.213 | 16350 | -0.909 | 1.803 | 785 |

* Notes: This table reports estimates of the effect of additional Internet on job outcomes and searches, as well as a placebo estimate of the common trends assumption following de Chaisemartin and D'Haultfœuille (2020). Standard errors are clustered by districts.

used as a proxy for skill level here.

$$\begin{aligned} Y_{ijt} = \alpha &+ \beta_1(PercentConnected_{jt} \times Primary_i) \\ &+ \beta_2 PercentConnected_{jt} + \beta_3 Primary_i + X'_{ijt}\delta + \gamma_t + \theta_j + \epsilon_{ijt} \end{aligned} \tag{7}$$

where the dummy variable $Primary_{it}$ indicates whether individual i has completed primary education at time t.

Table 7 shows the estimated results of how skilled and unskilled workers are impacted by Internet access. With additional Internet available, individuals who have finished primary education or more are 7.4 percentage points more likely to be employed than those who have no school or haven't finished primary school (column 1). Total income for the educated is higher as well (column 2). These differences are not statistically significant.

Column 3-7 report how their search behaviors differ given the Internet access change. Skilled workers use less numbers of different search methods, and are 11.8 percentage points more likely to use online searches compared to unskilled workers.

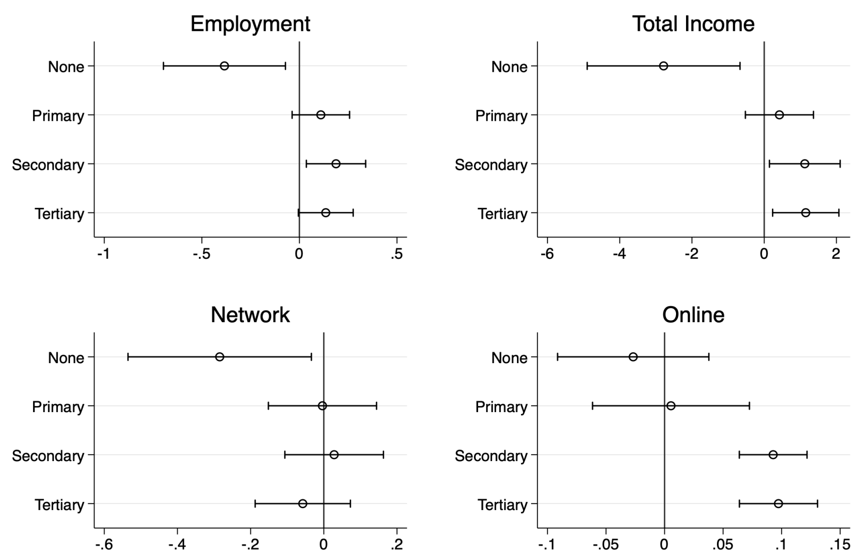
Table 7: Impacts of Internet Connection on Job Outcomes by Education

| Outcome | Employed | Income | No.of Methods | Online | | Network | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| % connected | 0.081 | 0.133 | -0.082 | -0.010 | -0.030 | -0.016 | 0.008 |
| | (0.075) | (0.453) | (0.061) | (0.034) | (0.043) | (0.074) | (0.075) |
| ... × primary | 0.074 | 0.999 | -0.044 | 0.118*** | 0.112** | -0.019 | 0.016 |
| | (0.071) | (0.624) | (0.075) | (0.038) | (0.045) | (0.049) | (0.049) |
| ... × parents primary | | | | | 0.122*** | | -0.283*** |
| | | | | | (0.045) | | (0.055) |
| primary | 0.163*** | 1.524*** | 0.186*** | 0.090*** | 0.099*** | 0.094*** | 0.079*** |
| | (0.010) | (0.101) | (0.017) | (0.006) | (0.007) | (0.010) | (0.011) |
| Mean of outcome | 0.367 | 2.356 | 0.258 | 0.062 | 0.067 | 0.248 | 0.265 |
| Observations | 36,892 | 32,278 | 37,485 | 32,724 | 21,335 | 32,725 | 21,336 |
| R-squared | 0.147 | 0.168 | 0.031 | 0.060 | 0.069 | 0.053 | 0.051 |
| Year FE | Y | Y | Y | Y | Y | Y | Y |
| Location FE | Y | Y | Y | Y | Y | Y | Y |
| Controls | Y | Y | Y | Y | Y | Y | Y |

* Notes: Only workers between age 15 and 65 are included. All specifications include year, location fixed effects, age and gender control variables. Standard errors (in parentheses) are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For unskilled workers, Internet access discourages the usage of online channel by 1 percentage point (column 4). The difference in using network searches between skilled and unskilled are small (column 6). Unskilled workers could have different social economic status and network quality, so I include additional controls for parents' education interacted with Internet access. The results are similar for online job search (column 4,5), but change quite a lot for network search (column 6,7), suggesting that existing social network variances can play a role in the choice of job search methods.

Figure 4: Internet effects by education levels




*Notes: Each panel plots coefficients on dummies for highest education level, including controls for age and gender, location and year fixed effects. 95% confidence intervals are displayed.

To test whether the effect of Internet on labor outcomes varies by years of education, Figure 4 shows estimation of equation 7 with dummies for no school, primary, secondary, and tertiary education levels.The results indicate that Internet connection increases the employment and income among the more educated workers the most. This finding is similar to Hjort and Poulsen (2019).

## 5.4 Heterogeneous effects by age

I also examine the heterogeneous effects of Internet exposure on job outcomes and search channels by workers' age, specifically young workers between 15 and 24 years old. Table 8 displays the estimation results of equation 8.

$$
\begin{aligned}
Y_{ijt} = \alpha + \beta_1(PercentConnected_{jt} \times Youth_i) \\
+ \beta_2 PercentConnected_{jt} + \beta_3 Youth_i + X'_{ijt}\delta + \gamma_t + \theta_j + \epsilon_{ijt}
\end{aligned} \quad (8)
$$

where the dummy variable $Youth_{it}$ indicates whether individual i is between 15 and 24 years old at time t.

Compared with experienced workers in areas with more Internet, young workers are significantly paid less (column 1, 2). Young workers will try more number of search methods (column 3), and more likely to increase searching through personal networks (column 6,7). No significant difference are found for online job search (column 4,5). These results suggest that young workers spend more effort searching for jobs, but the methods they choose are not as effective as the experienced. Their increasing reliance on personal networks suggests that Internet could also make it easier to communicate with family and friends using tools such as emails or social media.

Table 8: Impacts of Internet Connection on Job Outcomes by Age

| Outcome | Employed | Income | No.of Methods | Online | | Network | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| % connected | 0.141* | 1.330*** | -0.196*** | 0.065*** | 0.045** | -0.035 | -0.017 |
| | (0.073) | (0.392) | (0.060) | (0.013) | (0.021) | (0.057) | (0.061) |
| ... × youth | 0.007 | -1.376** | 0.332*** | -0.003 | -0.009 | 0.062 | 0.149*** |
| | (0.055) | (0.625) | (0.103) | (0.025) | (0.034) | (0.045) | (0.043) |
| ... × parents primary | | | | | 0.126*** | | -0.241*** |
| | | | | | (0.046) | | (0.048) |
| youth | -0.354*** | -2.599*** | -0.229*** | -0.056*** | -0.058*** | -0.282*** | -0.289*** |
| | (0.013) | (0.143) | (0.023) | (0.006) | (0.007) | (0.011) | (0.015) |
| Mean of outcome | 0.366 | 2.340 | 0.251 | 0.061 | 0.067 | 0.247 | 0.265 |
| Mean of youth | 0.144 | 0.741 | 0.244 | 0.035 | 0.038 | 0.132 | 0.148 |
| Observations | 37,034 | 32,411 | 38,436 | 32,855 | 21,300 | 32,856 | 21,301 |
| R-squared | 0.209 | 0.237 | 0.037 | 0.090 | 0.105 | 0.090 | 0.086 |
| Year FE | Y | Y | Y | Y | Y | Y | Y |
| Location FE | Y | Y | Y | Y | Y | Y | Y |
| Controls | Y | Y | Y | Y | Y | Y | Y |

[*] Notes: Only workers between age 15 and 65 are included. All specifications include year, location fixed effects, age, gender and education control variables. Standard errors (in parentheses) are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# 6 Mechanism Evidence

In this section, I discuss how Internet might affect employment and job search behaviors by skill levels and workers' age.

## 6.1 Access constraints

In Table 4, I find more Internet access enhances the use of online information search, and has little impact on the use of social network search. In section 5.3 and 5.4, I also show choices of search channel made by uneducated or young workers respond to more Internet availability differently than their peers. Even after Internet is made more available in their areas, individuals without primary education will not use online search, and young workers will increase their reliance on personal networks.

One possible explanation is that there are other constraints prevent disadvantaged workers from accessing the Internet for online job search. Figure A3 shows high cost of equipment is the most important reason for not having Internet access at home, according to the General Household Survey (GHS) in 2018.

If we consider the computer a tool necessary for online job search[3], computer ownership can be used as a proxy to test if accessing costs are different for heterogeneous workers. I use both the individual and household survey data from NIDS, and show test results in Table 9, Table 10. All regressions include individual fixed effects in additional to location and year fixed effects.

For skilled workers, their probability of owning a computer is 9 percentage point higher than unskilled workers, and they are more likely to be computer literate (Table 10,column 1-2). Their households are also more likely to spend money on Internet (column 5). Interestingly, I find cellphone ownership are lower for the skilled workers than the unskilled (column 3, 6). Young workers are obviously more tech-savvy: more likely to own a computer or cellphone, and know how to use a computer (Table 10,column 1-3). However, it seems that they are not using this technology to search for job information online directly, but rather to enhance personal networks for sharing job information. Most people probably communicate with families and friends or use social networking websites through a cellphone, so the more widely available cellphone could suggest no significant cost difference in accessing the Internet for network job search. This could explain why we do not see large differences in using network search for the skilled and unskilled in column 6 Table 7.

The findings about computer ownership and literacy suggest that technology will not make a difference if there are other constraints. Similar result are found in rural South Africa, where the rollout of mobile phone networks increased employment among women, but only for those who did not have significant family responsibilities (Klonner and Nolen, 2008).

[3] Smartphones which cost less can be a substitute for computers for many functions. However,

Table 9: Impacts of Internet Connection on Cell and PC Ownership by Education

| | Individual | | | Household | | | |
|---|---|---|---|---|---|---|---|
| | own a computer | computer literate | own a cellphone | own a computer | spent money on internet | own a cellphone | spent money on cellphone |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| % connected | -0.005 | -0.179*** | -0.012 | 0.023 | 0.005 | 0.176** | 0.054 |
| | (0.014) | (0.050) | (0.105) | (0.032) | (0.015) | (0.075) | (0.165) |
| ... × primary | 0.090*** | 0.256*** | -0.157*** | 0.107 | 0.034* | -0.228*** | -0.236* |
| | (0.026) | (0.061) | (0.054) | (0.071) | (0.020) | (0.047) | (0.124) |
| primary | -0.008 | 0.080*** | 0.182*** | -0.013* | -0.001 | 0.025** | 0.042** |
| | (0.005) | (0.015) | (0.016) | (0.007) | (0.004) | (0.011) | (0.019) |
| Mean of outcome | 0.051 | 0.293 | 0.714 | 0.107 | 0.012 | 0.852 | 0.740 |
| Observations | 33,829 | 32,850 | 33,841 | 35,715 | 26,777 | 35,768 | 26,724 |
| R-squared | 0.530 | 0.670 | 0.493 | 0.583 | 0.444 | 0.384 | 0.446 |
| Individual FE | Y | Y | Y | Y | Y | Y | Y |

* Notes: Only workers between age 15 and 65 are included. All specifications include individual, location and year fixed effects. Standard errors (in parentheses) are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in a survey of people who used smartphones to apply for a job, 47% had difficulties accessing content that did not display properly, 38% had difficulties entering in a large amount of text, 37% had difficulties submitting required files and supporting documentation, and 23% had difficulties bookmarking saved job applications for later( Smith 2015)

Table 10: Impacts of Internet Connection on Cell and PC Ownership by Age

| | Individual | | | Household | | | |
|---|---|---|---|---|---|---|---|
| | own a computer | computer literate | own a cellphone | own a computer | spent money on internet | own a cellphone | spent money on cellphone |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| % connected | 0.034 | -0.057 | -0.155* | 0.079 | 0.030 | 0.031 | -0.108 |
| | (0.021) | (0.044) | (0.085) | (0.052) | (0.023) | (0.064) | (0.113) |
| ... × youth | 0.093*** | 0.209*** | 0.304*** | 0.030 | -0.010 | 0.020 | 0.086 |
| | (0.029) | (0.060) | (0.082) | (0.028) | (0.020) | (0.037) | (0.096) |
| youth | -0.025*** | -0.089*** | -0.050*** | -0.010 | -0.001 | 0.003 | -0.034** |
| | (0.007) | (0.017) | (0.018) | (0.007) | (0.004) | (0.013) | (0.017) |
| Mean of outcome | 0.051 | 0.292 | 0.714 | 0.107 | 0.012 | 0.851 | 0.738 |
| Observations | 33,767 | 32,788 | 33,779 | 36,696 | 27,629 | 36,750 | 27,580 |
| R-squared | 0.532 | 0.671 | 0.492 | 0.579 | 0.436 | 0.380 | 0.440 |
| Individual FE | Y | Y | Y | Y | Y | Y | Y |

* Notes: Only workers between age 15 and 65 are included. All specifications include individual, location and year fixed effects. Standard errors (in parentheses) are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### 6.2 Internet activities - social networking

Job search response could also depend on the variety of uses of Internet technology. Table A1 Panel B shows that social networking is the most important Internet activity (44.5%), while only about 12% survey respondent uses Internet for job search [4]. It is possible that more Internet access not only reduces the cost of acquiring job information online, but also communication cost among family and friends, e.g. email, social media (Armona, 2021; Gee, Jones and Burke, 2017). I show that with Internet, young people increasingly rely on personal networks for information in Table 8. Whether the Internet causes individuals to search more or less using one method depends on the relative price changes of the different job search methods (Stevenson, 2008).

[4]Source: Research ICT Africa (RIA)

Using favor exchanges with people outside of the household in the past year as a dependent variable related to referral networks, Table 11 shows that with more Internet, people are less likely to exchange favors but the number of favors could increase.

Table 11: Impacts of Internet Connection on Favor Exchanges

| | Interactions with people outside of the household last year | | | | | |
|---|---|---|---|---|---|---|
| | exchanged favors | | given favors | | received favors | |
| | yes | no.of | yes | no.of | yes | no.of |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| % connected | -0.100 | 0.448 | -0.082 | 1.046** | -0.027 | -0.168 |
| | (0.070) | (0.500) | (0.064) | (0.500) | (0.037) | (0.967) |
| Mean of outcome | 0.138 | 2.293 | 0.055 | 2.298 | 0.090 | 2.303 |
| Observations | 34,534 | 1,884 | 34,213 | 538 | 34,533 | 1,029 |
| R-squared | 0.369 | 0.541 | 0.362 | 0.567 | 0.365 | 0.567 |
| Individual FE | Y | Y | Y | Y | Y | Y |

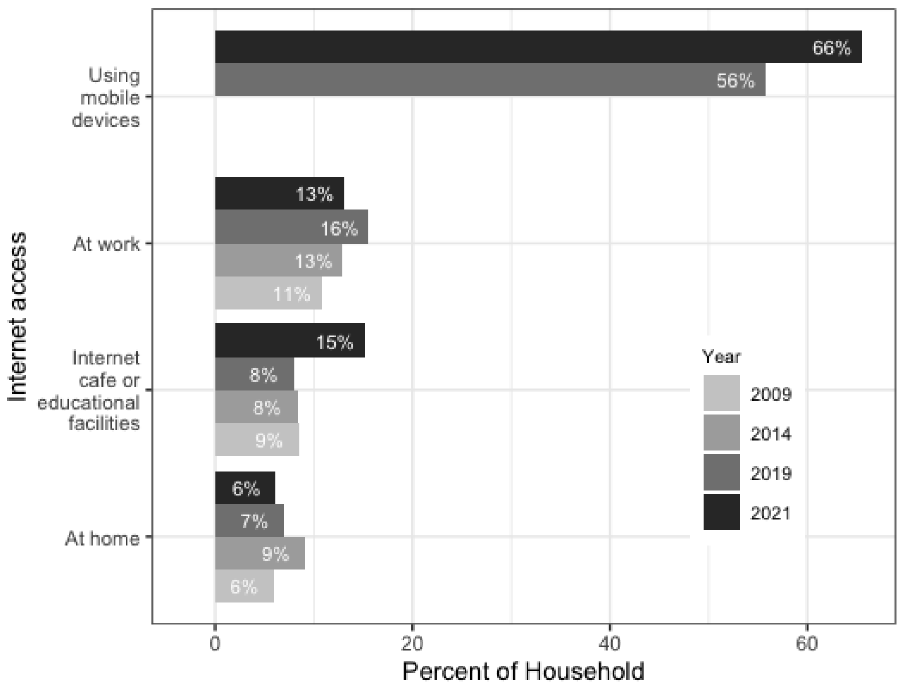
* Notes: Only workers between age 15 and 65 are included. All specifications include individual, location and year fixed effects. Standard errors (in parentheses) are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## 6.3 Adoptions by household and firms

In previous analysis, the key variable of interest "% connected" represents the Internet penetration rate in the district where the worker is living, it does not directly reflect the actual Internet access of the households or individual. As a supplement source, I use the General Household Survey (GHS) by Statistics South Africa to show some first stage correlations between Internet availability and adoption.

Households in South Africa are generally more likely to have access to the Internet at work than at home or at Internet cafes or at educational institutions. In 2021, Internet access using mobile devices (66%) is the most common way compared to access at home (6%), at work (16%) and elsewhere (15%) (Figure 5). I estimated these statistics using the GHS. The target population of the survey consists of all private households in all nine provinces of South Africa and residents in workers' hostels. The sample size is about 24 thousand households each year, from 2009 to 2021. This survey includes information about whether households had at least one member who had access to, or used the Internet either at home, work, place of study or Internet cafes, which can be used as proxy for direct Internet adoption rate.

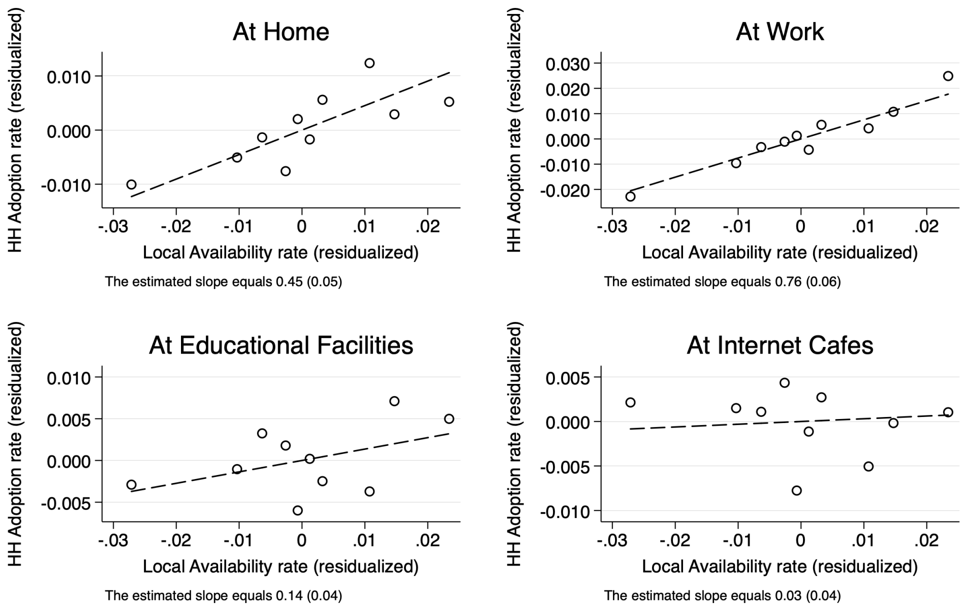
Figure 5: Households' Internet Access by Place of Access




Source:General Household Survey, Statistics South Africa.

I find that relationships between broadband Internet availability in the local area and households' Internet adoption rates vary by places of access. Internet access types data from the GHS can be matched with the percent populations connected "% connected" data by province and year. Figure 6 shows scatterplots of the Internet adoption rate against the Internet availability rate in the province, after taking out provincial and year fixed effects. The x-axis reports residuals from a regression of percent of populations connected to broadband on province and year fixed effects, and the y-axis reports residuals from a regression of households having Internet by access types on province and year fixed effects.

Figure 6: Internet availability and household adoption rate, by places of access




Note: The scatter plot shows average (residual) adoption at (residual) availability deciles, by places of access

The figure is based on the following regression that uses the sample of households for which we observe whether or not they have Internet access at home, work, nearby Internet cafes, or educational facilities:

$$d_{ijt} = \delta PercentConnected_{jt} + \gamma_t + \theta_j + \nu_{ijt} \tag{9}$$

where $d_{ijt}$ equals one if household i in province j at time t had Internet access at home (or at work, at educational facilities, at nearby Internet cafe) and is zero otherwise.

The coefficient on the availability rate $\delta$ is about 0.45 with a standard error of 0.05 for Internet access at home. This estimate implies that a 10 percentage point increase in broadband availability will induce 4.5 percent of households to gain Internet access at home. Adoption at work responses the most, while access from

Internet cafes do not change much. These findings illustrate that when Internet becomes available, adoption is not random; instead, it is more likely adopted in places in which complementary factors are abundant, including computers and computer literacy.

Table 12: Impacts of Internet Connection on Adoptions by Place of Access

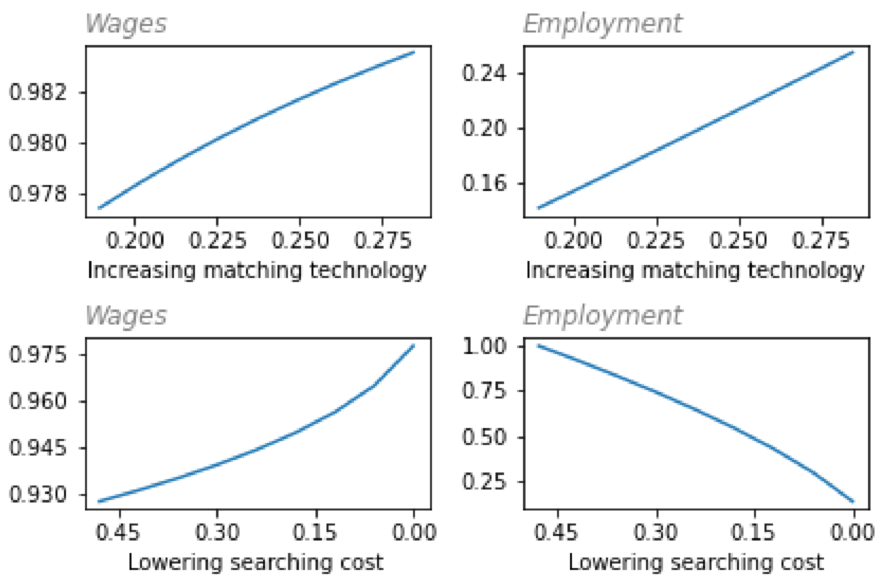
| | Anywhere | At home | At work | Educational facilities | Internet cafes |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| % connected | 0.946*** | 0.427*** | 0.805*** | 0.158* | 0.086 |
| | (0.173) | (0.073) | (0.113) | (0.090) | (0.135) |
| Mean of outcome | 0.220 | 0.075 | 0.133 | 0.058 | 0.048 |
| Observations | 127,024 | 126,349 | 126,349 | 127,024 | 127,024 |
| R-squared | 0.106 | 0.046 | 0.069 | 0.033 | 0.029 |

* Notes: All specifications include location and year fixed effects. Standard errors (in parentheses) are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## 6.4 General equilibrium impacts

Internet can affect both the labor supply and the labor demand, and the results on employment and wages in Table 3 and Table 7 should reflect the equilibrium outcomes. Without employers or firms' data, I can not say much about how employers' job creation decisions respond to more Internet access empirically. Instead, I study the expected general equilibrium impacts of Internet using the standard theory of unemployment and vacancies , the Diamond, Mortensen and Pissarides (DMP) Model (Mortensen, 2000). Simulated equilibrium changes of wage and employment are summarized in Figure 7. A brief description of the DMP framework is presented in Appendix B.

Figure 7: New equilibrium simulation using DMP model




Notes: By changing the parameter value of matching technology $A_t$ and the value of unemployment income $b$, I numerically solve the new equilibrium after a Internet access shock. The baseline parameter values are from Hagedorn and Manovskii (2008).

The first mechanism that Internet could impact the labor market is an improvement in matching efficiency. A key process in the DMP-framework is the "matching function", which uses job vacancies and job seekers as input, and produces a number of firm-worker matches given a matching technology $A$. Upper panel, Figure 7 show that with higher values of $A$, more hires can be generated from the same number of job seekers and vacancy, thereby increasing the employment rate. Since job seekers expect to be matched faster, their outside option improves, which will drive up wages in new employment relationships.

Internet access can also reduces the cost of learning about and applying for jobs. Unemployment income in the DMP model include both actual unemployment transfer and imputed value of time to unemployed workers. If we assume the search effort is fixed given the Internet, lower searching cost implies higher value of leisure, thereby increasing the value of unemployment income. Everything else equal, this increasing unemployment benefits exerts an upward pressure on the equilibrium wage. This lowers the profits employers receive from filled jobs, leading to a decline in vacancy creation. Lower vacancies imply a lower job finding rate for workers, which leads to an decrease in employment as shown in lower panel, Figure 7.

Combining these two mechanisms, the effect on wages is unambiguously positive, but the total effect of the Internet on employment depends on the relative importance of these two.

# 7 Conclusion

This paper provides evidence on how Internet availability affects job market outcomes and job search activity in South Africa. By comparing individuals in areas with various Internet penetration rates, I find that jobseekers in locations with better connectivity have higher employment rates and income, and the impact is driven by a significant increase in employment of experienced and skilled workers. When Internet is made more available, only skilled ones increase their use of online job search. Young workers will search through more methods, while rely more on personal networks.

These findings suggest that not everyone stands to benefit from improved Internet availability automatically. Associated labor market disruptions can be painful and can result in higher inequality. High cost remains the largest barrier for Internet usage [5]. The low-skilled or less-educated almost exclusively use mobile phones to access the Internet. Poor computer literacy could limit the productive use of this technology. Besides improving Internet infrastructure, complementary policies aiming at updating skill and digital literacy are critical for ensuring the overall benefits be shared broadly.

[5] Table A1 Africa ICT access survey, Fig A3

# References

**Abebe, Girum, A Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, and Simon Quinn.** 2021. "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City." *The Review of Economic Studies*, 88(3): 1279–1310.

**Abel, Martin, Rulof Burger, and Patrizio Piraino.** 2020. "The Value of Reference Letters: Experimental Evidence from South Africa." *American Economic Journal: Applied Economics*, 12(3): 40–71.

**Abel, Martin, Rulof Burger, Eliana Carranza, and Patrizio Piraino.** 2019. "Bridging the Intention-Behavior Gap? The Effect of Plan-Making Prompts on Job Search and Employment." *American Economic Journal: Applied Economics*, 11(2): 284–301.

**Aker, Jenny C.** 2010. "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger." *American Economic Journal: Applied Economics*, 2(3): 46–59.

**Aker, Jenny C, and Isaac M Mbiti.** 2010. "Mobile Phones and Economic Development in Africa." *Journal of Economic Perspectives*, 24(3): 207–232.

**Akerman, Anders, Ingvil Gaarder, and Magne Mogstad.** 2015. "The Skill Complementarity of Broadband Internet." *The Quarterly Journal of Economics*, 130(4): 1781–1824.

**Arbex, Marcelo, Dennis O'Dea, and David Wiczer.** 2019. "Network Search: Climbing the Job Ladder Faster." *International Economic Review*, 60(2): 693–720.

**Armona, Luis.** 2021. "Online Social Network Effects in Labor Markets: Evidence From Facebook's Entry to College Campuses." *SSRN Electronic Journal*, 75.

**Atasoy, Hilal.** 2013. "The Effects of Broadband Internet Expansion on Labor Market Outcomes." *ILR REVIEW*, 33.

**Autor, David H.** 2001. "Wiring the Labor Market." *Journal of Economic Perspectives*, 16.

**Banerjee, Abhijit, and Gaurav Chiplunkar.** 2020. "How Important Are Matching Frictions in the Labor Market? Experimental & Non-Experimental Evidence from a Large Indian Firm." 52.

**Beaman, Lori, and Jeremy Magruder.** 2012. "Who Gets the Job Referral? Evidence from a Social Networks Experiment." *American Economic Review*, 102(7): 3574–3593.

**Beaman, Lori, Niall Keleher, and Jeremy Magruder.** 2018. "Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi." *Journal of Labor Economics*, 36(1): 121–157.

**Bhuller, Manudeep, Andreas Kostol, and Trond C. Vigtel.** 2019. "How Broadband Internet Affects Labor Market Matching." *SSRN Electronic Journal*.

**Bonnet, Florence, Joann Vanek, and Martha Chen.** 2019. "Women and Men in the Informal Economy – A Statistical Brief." International Labour Office, Geneva.

**Borusyak, Kirill, Xavier Jaravel, and Jann Spiess.** 2022. "Revisiting Event Study Designs: Robust and Efficient Estimation."

**Callaway, Brantly, and Pedro H.C. Sant'Anna.** 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225(2): 200–230.

**Calvó-Armengol, Antoni, and Matthew O. Jackson.** 2004. "The Effects of Social Networks on Employment and Inequality." *The American Economic Review*, 94(3): 426–454.

**Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller.** 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics*, 90(3): 414–427.

**Card, David, Jochen Kluve, and Andrea Weber.** 2018. "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." *Journal of the European Economic Association*, 16(3): 894–931.

**Caria, Stefano, Theres Lessing, and Ann Hermes.** 2019. "Filling the Gap: How Information Can Help Jobseekers."

**Carranza, Eliana, Robert Garlick, and Kate Orkin.** 2020. "Job Search and Matching with Two-Sided Limited Information."

**Casella, Alessandra, and Nobuyuki Hanaki.** 2008. "Information Channels in Labor Markets: On the Resilience of Referral Hiring." *Journal of Economic Behavior & Organization*, 66(3-4): 492–513.

**de Chaisemartin, Clément, and Xavier D'Haultfœuille.** 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996.

**de Chaisemartin, Clément, and Xavier D'Haultfœuille.** 2022. "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey." NBER Working Paper 29691.

**de Chaisemartin, Clément, Xavier D'Haultfœuille, and Yannick Guyonvarch.** 2019. "Fuzzy Differences-in-Differences with Stata." *The Stata Journal: Promoting communications on statistics and Stata*, 19(2): 435–458.

**de Chaisemartin, Clément, Xavier D'Haultfoeuille, Félix Pasquier, and Gonzalo Vazquez-Bare.** 2022. "Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period."

**Dettling, Lisa.** 2017. "Broadband and the Labor Market: The Impact of Redidential High-speed Internet on Married Women's Labor Force Participation." *ILR Review*, 70(2): 451–482.

**Gee, Laura K., Jason Jones, and Moira Burke.** 2017. "Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network." *Journal of Labor Economics*, 35(2): 485–518.

**Goodman-Bacon, Andrew.** 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277.

**Goyal, Aparajita.** 2010. "Information, Direct Access to Farmers, and Rural Market Performance in Central India." *American Economic Journal: Applied Economics*, 2(3): 22–45.

**Granovetter, Mark S.** 1973. "The Strength of Weak Ties." *American Journal of Sociology*, 78(6): 1360–1380.

**Hagedorn, Marcus, and Iourii Manovskii.** 2008. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited." *American Economic Review*, 98(4): 1692–1706.

**Hagedorn, Marcus, Fatih Karahan, Iourii Manovskii, and Kurt Mitman.** 2013. "Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects." *National Bureau of Economic Research*, 76.

**Heath, Rachel.** 2018. "Why Do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories." *Journal of Political Economy*, 126(4): 1691–1746.

**Hjort, Jonas, and Jonas Poulsen.** 2019. "The Arrival of Fast Internet and Employment in Africa." *American Economic Review*, 109(3): 1032–1079.

**Holzer, Harry J.** 1988. "Search Method Use by Unemployed Youth." *Journal of Labor Economics*, 6(1): 1–20.

**Jack, William, and Tavneet Suri.** 2014. "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution." *American Economic Review*, 104(1): 183–223.

**Jensen, Robert.** 2007. "The Digital Provide: Information (Technology), Market Performance, and Welfare in The." *Source: The Quarterly Journal of Economics*, 122(3): 879–924.

**Klonner, Stefan, and Patrick Nolen.** 2008. "Does ICT Bene...t the Poor? Evidence from South Africa." 37.

**Kroft, Kory, and Devin G. Pope.** 2014. "Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist." *Journal of Labor Economics*, 32(2): 259–303.

**Kuhn, Peter, and Hani Mansour.** 2014. "Is Internet Job Search Still Ineffective?" *The Economic Journal*, 124(581): 1213–1233.

**Kuhn, Peter, and Mikal Skuterud.** 2004. "Internet Job Search and Unemployment Durations." *American Economic Review*, 94(1): 218–232.

**La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea.** 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics*, 4(4): 1–31.

**Long, Wenjin, Simon Appleton, and Lina Song.** 2017. "The Impact of Job Contact Networks on Wages of Rural–Urban Migrants in China: A Switching Regression Approach." *Journal of Chinese Economic and Business Studies*, 15(1): 81–101.

**Michaels, Guy, Ashwini Natraj, and John Van Reenen.** 2014. "Has ICT Polarized Skill Demand? Evidence from Eleven Countries Over Twenty-five Years." *The Review of Economics and Statistics*, 96(1): 60–77.

**Montgomery, James D.** 1991. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis." *The American Economic Review*, 81(5): 1408–1418.

**Mortensen, Dale T.** 1986. "Chapter 15 Job Search and Labor Market Analysis." In *Handbook of Labor Economics*. Vol. 2, 849–919. Elsevier.

**Mortensen, Dale T.** 2000. "Modeling Matched Job-Worker Flows." Econometric Society 1493.

**Mortensen, Dale T., and Christopher A. Pissarides.** 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *The Review of Economic Studies*, 61(3): 397–415.

**Mouw, Ted.** 2003. "Social Capital and Finding a Job: Do Contacts Matter?" *American Sociological Review*, 68(6): 868–898.

**Pellizzari, Michele.** 2010. "Do Friends and Relatives Really Help in Getting A Good Job?" *Industrial and Labor Relations Review*, 63(3): 494–510.

**Smith, Aaron.** 2015. "Searching for Work in the Digital Era." Pew Research Center.

**Stevenson, Betsey.** 2008. "The Internet and Job Search." National Bureau of Economic Research, Inc NBER Working Paper 13886.

**Sun, Liyang, and Sarah Abraham.** 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199.

**Topa, Giorgio.** 2011. "Labor Markets and Referrals." In *Handbook of Social Economics.* Vol. 1, 1193–1221. Elsevier.

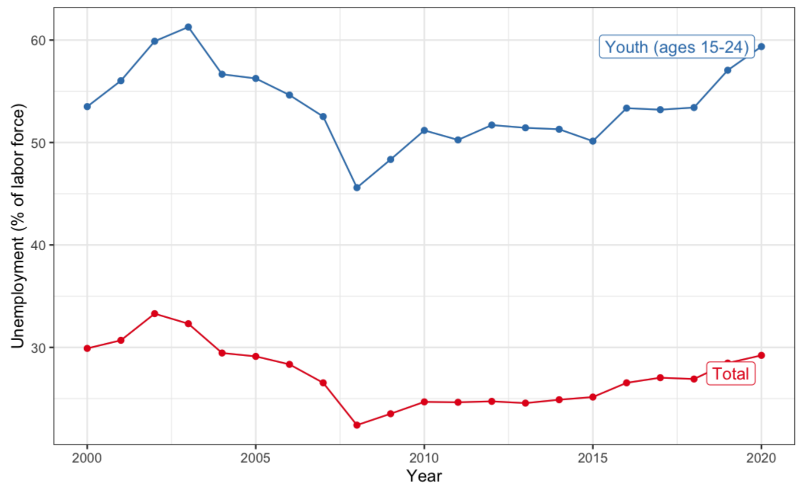
**Wheeler, Laurel E, Robert Garlick, Eric Johnson, Patrick Shaw, and Marissa Gargano.** 2022. "Linkedin(to) Job Opportunities: Experimental Evidence from Job Readiness Training." *American Economic Journal: Applied Economics.*

**Wolpin, Kenneth I.** 1987. "Estimating a Structural Search Model: The Transition from School to Work." *Econometrica*, 55(4): 801–817.

**World Development Report 2016.** n.d.. "World Development Report 2016." *world development report*, 359.

# Appendices

Figure A1: Unemployment Rate in South Africa




Source: World Development Indicators.

Figure A2: Employment before and after Internet arrival by penetration level

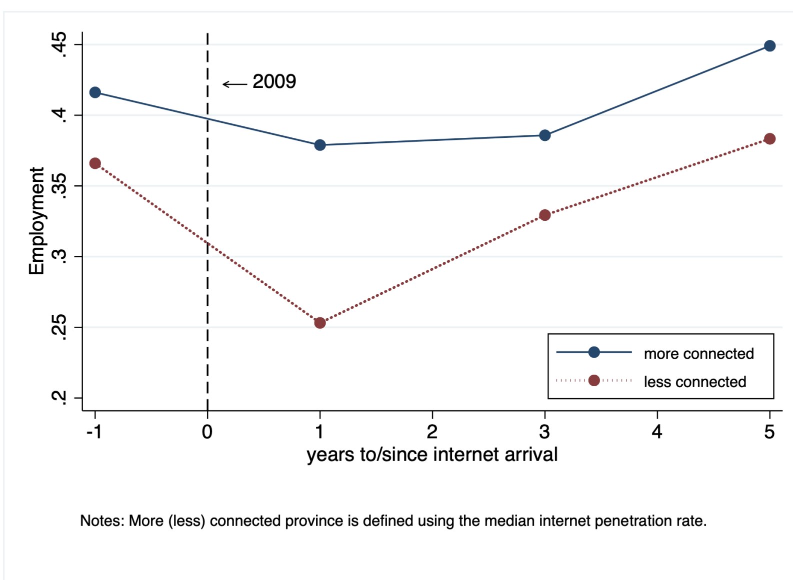
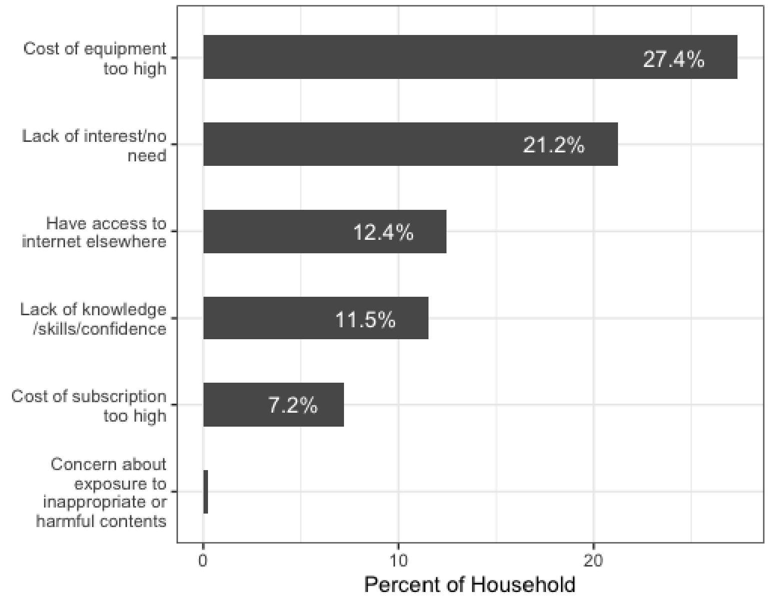


Table A1: South Africa ICT access survey

| | 2017-2018 | 2011-2012 | 2005-2008 |
|---|---|---|---|
| **Panel A: household attributes** | | | |
| *HH has internet connection* | 11.6% | 16.2% | 6.5% |
| *HH with Internet: highest education level* | | | |
| No school | 0.9% | | |
| Primary | 1.4% | | |
| Secondary and above | 97.6% | | |
| *Reasons not having internet* | | | |
| Cost too high | 48.3% | | |
| Not available in the area | 5.9% | | |
| Do not need | 20.4% | | |
| Do not know how to use it | 12.4% | | |
| Others | 12.9% | | |
| **Panel B: Individual usage** | | | |
| *Used Internet before* | 68.6% | 33.0% | 18.6% |
| *Internet usage* | | | |
| Once a day | 50.4% | 64.8% | 64.4% |
| Once a week | 30.8% | 24.6% | 24.9% |
| Once a month | 10.3% | 9.1% | 7.0% |
| Less than once a month | 8.6% | 1.5% | 3.6% |
| *Most important internet activity* | | | |
| Social networking | 44.5% | | |
| Education | 23.5% | | |
| Job search | 12.4% | | |
| Work related | 11.3% | | |
| Online banking | 2.5% | | |
| Others | 5.7% | | |
| *Limitation for use of the internet (multiple responses)* | | | |
| Cost | 46.6% | 62.9% | 45.3% |
| Speed | 25.6% | 10.1% | 8.8% |
| No interesting content in my language | 7.1% | | 13.0% |
| Difficult to use | 2.8% | 73.2% | 1.5% |
| *Reason not using internet(single choice)* | | | |
| Cost | 50.6% | | |
| No interest | 19.3% | | |
| Do not know how to use it | 8.9% | | |
| Not available in my area | 3.4% | | |
| Others | 17.9% | | |

Source: Africa ICT access survey.

Figure A3: Reasons for not having Internet access at home




Source: General Household Survey, 2018, Statistics South Africa.

# A A model of job seeker's utility maximization with leisure

I include leisure in the utility function for job seekers in this version of the conceptual model. A job seeker lives two periods with a supply of Internet access $\theta$. In the first period, an unemployed individual receives some unemployment benefit $b$, and needs to allocates his time (normalized to 1) between job searching $s$ and leisure $l$. The probability of finding a job depends on the search effort and amount of Internet access: $p(s, \theta)$. In the second period, if the individual becomes employed, the wage and labor supplied will be given as $w$ and $h$. The job seeker chooses job search effort and maximizes the expected lifetime utility as follows:

$$
\begin{aligned}
\max_{s} \quad & u\left(c_1,\ell_1,\theta\right)+\beta E u\left(c_2,\ell_2,\theta\right) \\
\text{s.t.} \quad & c_1 = b \\
& \ell_1 = 1-s \\
& c_2 = \begin{cases} wh & \text{w.p. } p(s,\theta) \\ b & \text{w.p. } 1-p(s,\theta) \end{cases} \\
& \ell_2 = \begin{cases} 1-h & \text{w.p. } p(s,\theta) \\ 1 & \text{w.p. } 1-p(s,\theta) \end{cases} \\
& 0 \le s, p(s,\theta) \le 1
\end{aligned}
\tag{A1}
$$

An interior solution should satisfy the following first order condition:

$$
\frac{\partial u\left(b,1-s,\theta\right)}{\partial l_1} = \beta\frac{\partial p(s,\theta)}{\partial s}\left[u\left(wh,1-h,\theta\right)-u\left(b,1,\theta\right)\right] \tag{A2}
$$

which implies that the individual chooses search effort $s$ optimally such that the marginal utility of giving up leisure equals the expected utility gain from searching for work, which is the difference between employment and unemployment utility in the second period.

For this paper, I am interested in how employment probability may change with the Internet access, which is provided exogenously. That is,

$$
\frac{d}{d\theta}p(s(\theta),\theta) = \frac{\partial p}{\partial s}s'(\theta)+\frac{\partial p}{\partial \theta} \tag{A3}
$$

Assuming the marginal productivity of search and Internet are both positive ($\frac{\partial p}{\partial s}, \frac{\partial p}{\partial \theta} > 0$), the effect on employment will depend on $s'(\theta)$. In order to see how optimal search effort $s^{\star}(\theta)$ changes with Internet access $\theta$, we can differentiate the first order condition equation A2 with respect to $\theta$:

$$
s'(\theta) = \frac{\beta p_{s\theta}\left(u^{emp}-u^{unemp}\right)+\beta p_s \frac{\partial}{\partial\theta}\left(u^{emp}-u^{unemp}\right)-u_{\ell\theta}}{-u^1_{\ell\ell}-\beta p_{ss}\left(u^{emp}-u^{unemp}\right)} \tag{A4}
$$

where $u^1, u^{emp}, u^{unemp}$ represent the utility in period 1, being employed and unemployed in period 2 respectively.

Since $u^{emp} > u^{unemp}$ is a necessary condition for the existence of an interior solution, the denominator in equation A4 is positive. The sign of the numerator depends on three parts. First, $p_{s\theta}$, the change in the marginal productivity of search in response to more Internet access. Second, $\frac{\partial}{\partial\theta}(u^{emp} - u^{unemp})$, the difference between employment and unemployment utility in response to more Internet access. Third, $u_{\ell\theta}$, the change in marginal utility from leisure in response to more Internet access.

# B DMP framework

I summarize the standard equilibrium search and matching model briefly in this appendix.

The hiring process is governed by a matching function that produces worker-employer pairs using job vacancies and job seekers as inputs,

$$H_t = A_t v_t^{\alpha} u_t^{1-\alpha} \tag{A5}$$

where $u_t$ is the number of job seekers, $v_t$ is the number of vacant jobs, and $A_t$ is the efficiency of the search and matching process.

The probability of finding a job match for the unemployed worker is given by $A_t(v_t/u_t)^{\alpha} = A_t(\theta_t)^{\alpha}$, where $\theta_t$ represents the labor market tightness.

All workers face the same constant unemployment risk $\lambda$. At steady states, the flow into unemployment $\lambda(1-u)$ should equal the flow out of unemployment $A\theta^{\alpha}u$. Unemployment can be solved in terms of two transition rates,

$$u = \frac{\lambda}{\lambda + A(\theta)^{\alpha}} \tag{A6}$$

Workers maximize the net present value of income and randomly search for vacant jobs while unemployed. The flow value of being unemployed is $rU = b + A(\theta)^{\alpha}(W - U)$, and the flow value of working is $rW = w + \lambda(U - W)$. Firms receive a flow value of profits for active jobs according to $rJ = p - w - \lambda J$, and the flow value of vacancy is $rV = -c + A(\theta)^{\alpha-1}(J - V)$. In profit-maximizing equilibrium, the expected value of a vacancy is driven to zero by free entry of new vacancies. We can derive the job creation condition as,

$$p - w - \frac{(r+\lambda)c}{A(\theta)^{\alpha}} = 0 \tag{A7}$$

The wage is assumed to be derived from a Nash bargaining solution: the $w$ that maximizes the weighted product of the worker's and the firm's net return from the job match.

$$w = \arg\max\,(W-U)^{\beta}\,(J-V)^{1-\beta}\,, \tag{A8}$$

where $\beta$ can interpreted as a relative measure of labor's bargaining strength, and it is between 0 and 1. First order condition gives the wage setting condition as,

$$w = (1-\beta)b + \beta p(1 + c\theta) \tag{A9}$$

Equilibrium is a unique set of $(u, \theta, w)$ that satisfies the flow equilibrium condition A6, the job creation condition A7, and the wage equation A9. By changing the parameter value of matching technology $A_t$ and the value of unemployment income $b$, I numerically solve the new equilibrium after a Internet access shock. The simulated results are shown in Figure 7.