Who benefits from better internet connectivity?

- evidence from the labor market in South Africa

Zhe Liu

Department of Economics, University of Washington

Northwest Development Workshop
May 2022

Overview

Motivation

- High level of unemployment in South Africa, especially among young workers. ▸ Unemployment rate
- Widely used network vs recent growth in online job search and hiring.
- Policy implications: To what extent, open and anonymous information source like online can substitute for exclusionary personal connections.
- Background: gradual arrival of the first fast undersea cable in Africa, which provided data transmission capacity to connected countries 10 times greater. ▸ Fast internet

Research Questions

1. Does the internet help improve the job outcomes (eg. income, employment) for workers with different skill levels?
2. How do job search channels (online and network) respond to faster and easier internet access?

Research Questions

1. Does the internet help improve the job outcomes (eg. income, employment) for workers with different skill levels?
2. How do job search channels (online and network) respond to faster and easier internet access?

Potential contributions:

- Internet's impacts on workers with different skill levels in developing countries
  - ICT in developing countries: focus on price variation: (Jensen 2007, Aker 2010, Goyal 2010), risk sharing (Jack and Suri 2014), cultural changes (La Ferrara, Chong, and Duryea 2012)
  - ICT in labor market: mixed empirical results in developed countries (Kroft and Pope 2010, Dettling 2017, Bhuller, Kostol and Vigtel 2019)
  - One direct evidence in developing countries - large positive in Africa (Hjort and Poulsen 2019)
- Evidence on search channels and workers' search behaviors

Main results preview

1. A large positive effects on employment rates and income for individuals with primary educations when internet arrives.

Main results preview

1. A large positive effects on employment rates and income for individuals with primary educations when internet arrives.
2. The network channel is quite resilient given the internet access shock: on average, the probabilities of using online and network job searches both increase;
    - For less-educated workers, they use more network and less online searches.

Conceptual model

A job seeker lives two periods with a supply of internet access θ. Assume their labor supply h is inelastic. The job seeker chooses job search effort s and maximizes the expected lifetime utility as follows:

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

We are interested in how employment probability may change with the internet access. That is,

$$\frac{d}{d\theta}p(s(\theta),\theta) = \frac{\partial p}{\partial s}s'(\theta) + \frac{\partial p}{\partial \theta} \tag{2}$$

Assume the marginal productivity of search and internet are both positive ($\frac{\partial p}{\partial s}, \frac{\partial p}{\partial \theta} > 0$)

- If search effort increases with internet access ($s'(\theta) > 0$), employment will increase as well
- If search effort declines with internet access ($s'(\theta) < 0$), the net impact on employment is unclear.

How optimal search effort $s^{\star}(\theta)$ changes with internet access θ:

$$s'(\theta) = \frac{\beta p_{s\theta}\left(u^{emp} - u^{unemp}\right) + \beta p_s \frac{\partial}{\partial \theta}\left(u^{emp} - u^{unemp}\right) - u_{\ell\theta}}{-u^1_{\ell\ell} - \beta p_{ss}\left(u^{emp} - u^{unemp}\right)} \quad (3)$$

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

How optimal search effort $s^{\star}(\theta)$ changes with internet access θ:

$$s'(\theta) = \frac{\beta p_{s\theta}\left(u^{emp} - u^{unemp}\right) + \beta p_s \frac{\partial}{\partial \theta}\left(u^{emp} - u^{unemp}\right) - u_{\ell\theta}}{-u^1_{\ell\ell} - \beta p_{ss}\left(u^{emp} - u^{unemp}\right)} \quad (3)$$

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

- the denominator is positive: $u^{emp} > u^{unemp}$ is a necessary condition for the existence of an interior solution

How optimal search effort $s^{\star}(\theta)$ changes with internet access θ:

$$s'(\theta) = \frac{\beta p_{s\theta}\left(u^{emp} - u^{unemp}\right) + \beta p_s \frac{\partial}{\partial \theta}\left(u^{emp} - u^{unemp}\right) - u_{\ell\theta}}{-u^1_{\ell\ell} - \beta p_{ss}\left(u^{emp} - u^{unemp}\right)} \quad (3)$$

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

- the denominator is positive: $u^{emp} > u^{unemp}$ is a necessary condition for the existence of an interior solution
- the sign of the numerator depends on three parts:
  - $p_{s\theta}$: the change in the marginal productivity of search in response to more internet access.

How optimal search effort $s^{\star}(\theta)$ changes with internet access θ:

$$s'(\theta) = \frac{\beta p_{s\theta}\left(u^{emp} - u^{unemp}\right) + \beta p_s \frac{\partial}{\partial \theta}\left(u^{emp} - u^{unemp}\right) - u_{\ell\theta}}{-u^1_{\ell\ell} - \beta p_{ss}\left(u^{emp} - u^{unemp}\right)} \quad (3)$$

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

- the denominator is positive: $u^{emp} > u^{unemp}$ is a necessary condition for the existence of an interior solution
- the sign of the numerator depends on three parts:
  - $p_{s\theta}$: the change in the marginal productivity of search in response to more internet access.
  - $\frac{\partial}{\partial \theta}\left(u^{emp} - u^{unemp}\right)$, the difference between employment and unemployment utility in response to more internet access.

How optimal search effort $s^{\star}(\theta)$ changes with internet access θ:

$$s'(\theta) = \frac{\beta p_{s\theta}\left(u^{emp} - u^{unemp}\right) + \beta p_s \frac{\partial}{\partial \theta}\left(u^{emp} - u^{unemp}\right) - u_{\ell\theta}}{-u^1_{\ell\ell} - \beta p_{ss}\left(u^{emp} - u^{unemp}\right)} \quad (3)$$

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

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

Positive correlation between internet usage and search effort/job offers

Suggestive evidence using published data from a field experiment that Abel et al. (2019) have done with South Africa youth.

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

	(1) Search Hours	(2) Applications	(3) Empl Responses	(4) Job Offers	(5) 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Data sources

Labor - National Income Dynamic Studies (NIDS)of South Africa

- Five waves of survey data (2008 - 2017, every two years)
- Nationally representative sample of over 28,000 individuals in 7,300 households across the country.
- Employment status, search channel used, job outcomes, occupation, education, province, and other characteristics.

Table 2: Summary Statistics

	Obs	Mean	SD
Individual characteristics			
Urban area	27,975	0.46	0.50
Age	28,008	33.87	14.47
Female	28,008	0.60	0.49
Youth	28,008	0.36	0.48
Finished primary school	27,021	0.47	0.50
Own a cellphone	25,355	0.68	0.47
Own a computer	25,342	0.04	0.20
Is computer literate	24,415	0.26	0.44
Labor market outcomes			
Employed	26,592	0.35	0.48
Employed with a salary job	25,698	0.25	0.43
Self employed	25,698	0.05	0.21
Monthly income	23,337	792.18	2152.90
Job has permanent duration	6,318	0.55	0.50
Weekly hours	7,147	38.86	18.01
Job search methods			
Network	28,008	0.19	0.40
Online	28,008	0.05	0.21
Government agencies	28,008	0.03	0.16
Others	28,008	0.12	0.33

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.

Data sources

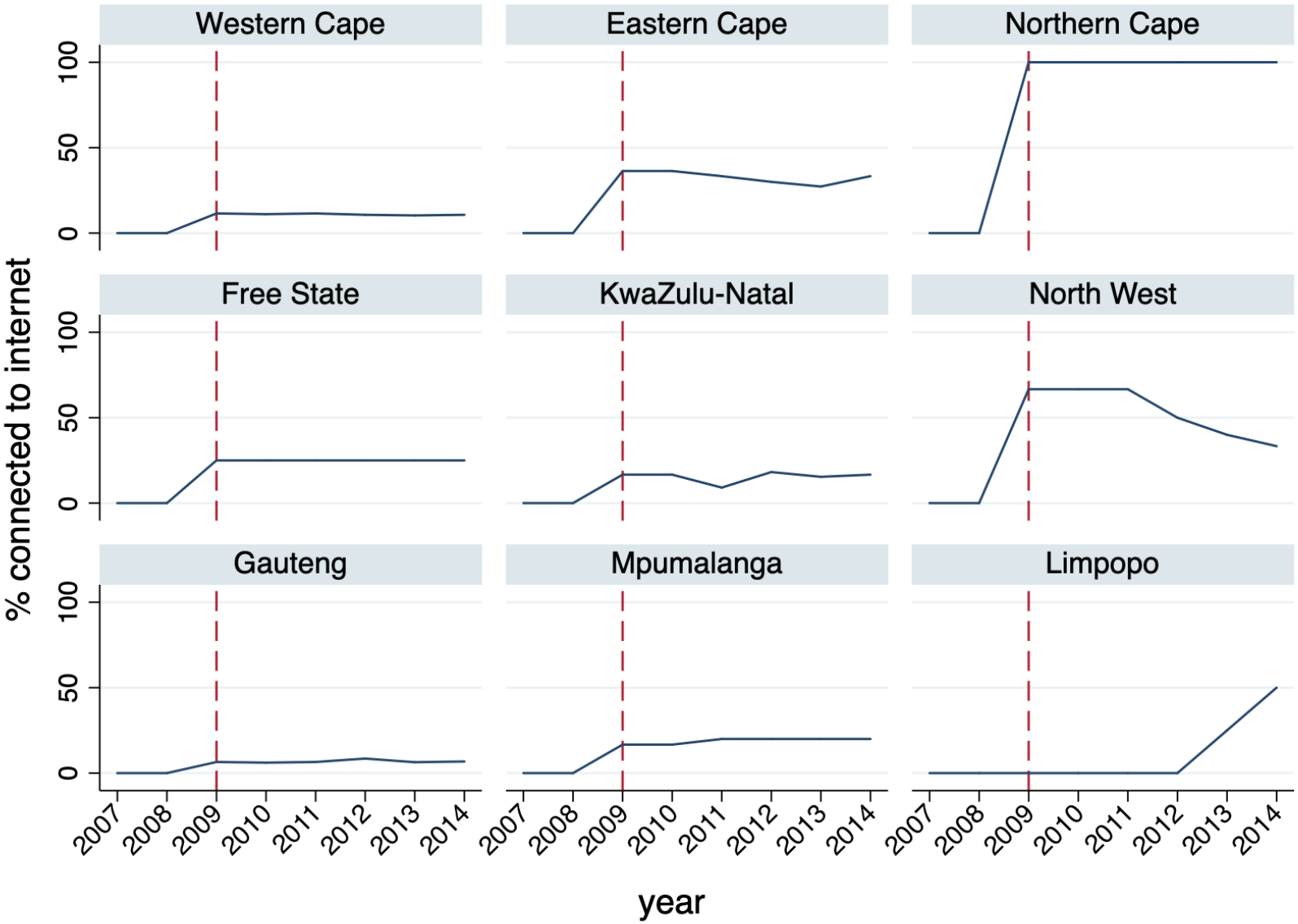
Internet access - Hjort & Poulsen (2019) AER paper

- Quarterly data on internet speeds, whether connected to the internet at city level from 2007 to 2014.

Matching two sources - using the geocode and year

- I aggregate the city-level connection data to province-level by calculating the percentage of cities with connection in one province by year.

Figure 1: Internet penetration level by provinces



Identification

- location fixed effects: controls for unobserved heterogeneity across provinces
- reverse causality: past levels of productivity variables (eg. employment, education, industry) does not predict internet connection

$$PercentConnected_{jt} = \alpha + X'_{ijt-1}\rho + \epsilon_{ijt} \quad (4)$$

Fixed effects estimation

- Average effect:

$$Y_{ijt} = \alpha + \beta PercentConnected_{jt} + \gamma_t + \theta_j + \epsilon_{ijt} \quad (5)$$

  - Y_{ijt} is the job outcomes or search channels for worker i in province j at time t
  - $PercentConnected_j$ is the percent of cities in province j connected to the backbone network
  - year fixed effects γ_t controls for nationwide, time-varying factors; province fixed effects θ_j controls for permanent, province-specific factors
  - X_{ijt} is a vector of individual- specific controls, including age and gender
  - β is the average effect of internet

- Heterogeneous effect

$$Y_{ijt} = \alpha + \beta_1(PercentConnected_{jt} \times Primary_i) + \beta_2 PercentConnected_{jt} + \beta_3 Primary_i + X'_{ijt}\delta + \gamma_t + \theta_j + \epsilon_{ijt} \quad (6)$$

  - $Primary_{it}$ is a dummy variable, indicating whether individual i has completed primary education at time t. A proxy for skill level
  - β_1 is the differential effect of internet on educated/skilled workers

Average effects

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

	(1) Employed	(2) Total income	(3) Online	(4) Network	(5) Num of methods
% connected	0.035	0.229	0.007	0.064*	0.031
	(0.031)	(0.202)	(0.020)	(0.034)	(0.094)
Mean of outcome	0.351	2.175	0.048	0.201	0.249
Observations	26,582	23,329	27,043	27,043	27,043
R-squared	0.100	0.105	0.494	0.468	0.451
Individual FE			Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Location FE	Y	Y	Y	Y	Y
Controls	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. Total income are calculated using monthly income if salary employed, profit if self-employed, and as zero if unemployed. Inverse hyperbolic sign are done to total income. Standard errors (in parentheses) are clustered at the province level. Control variables include age and gender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

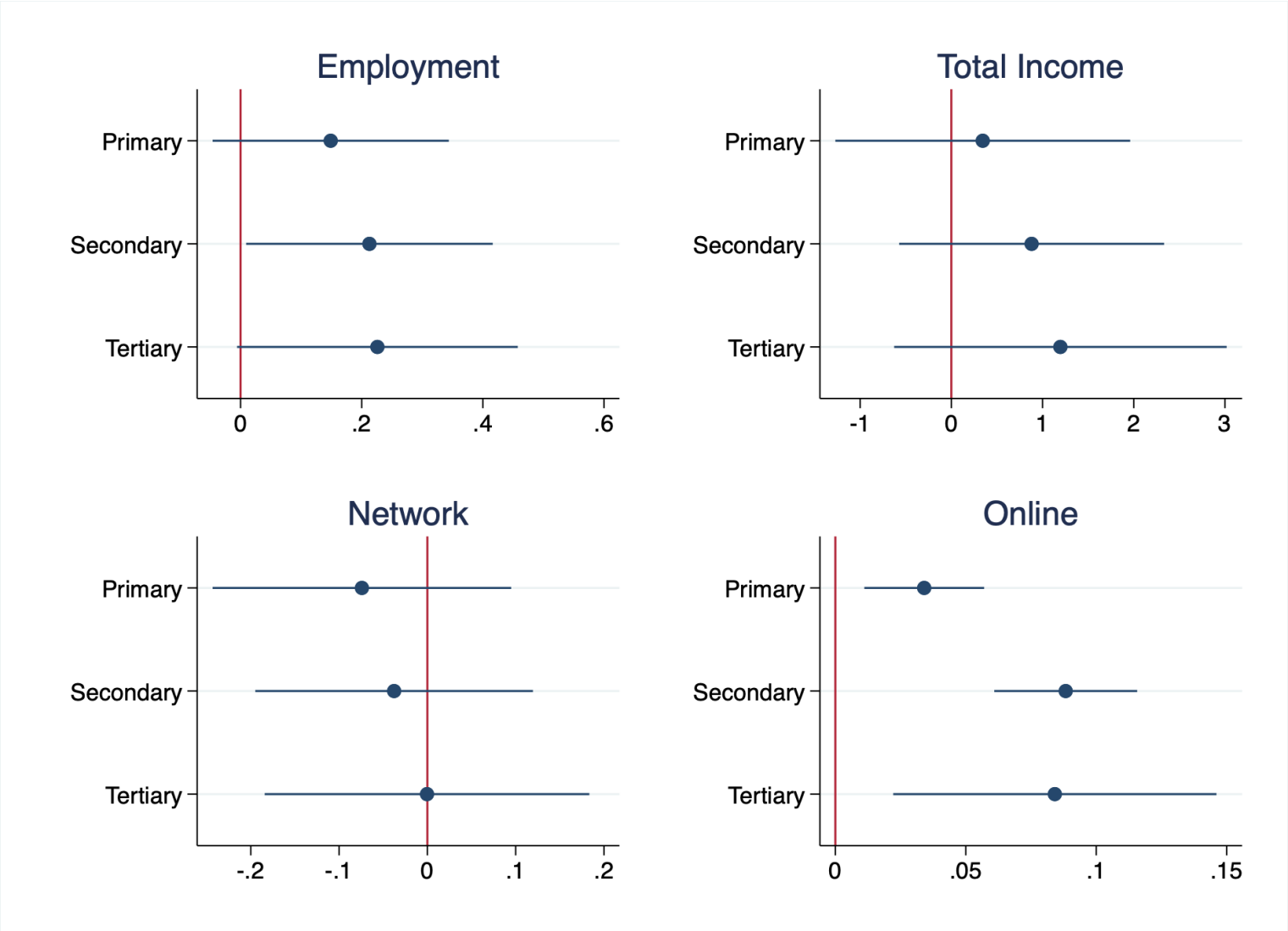
Heterogeneous effects

Table 4: Impacts of Internet Connection on Job Outcomes by Education

	(1) Employed	(2) Total income	(3) Online	(4) Network	(5) Num of methods
% connected	0.001	-0.075	-0.022	0.054**	-0.011
	(0.033)	(0.225)	(0.013)	(0.026)	(0.061)
... × Beyond primary	0.076**	0.656***	0.059***	0.018	0.081
	(0.030)	(0.215)	(0.021)	(0.029)	(0.060)
Beyond primary	0.155***	1.407***	0.074***	0.070***	0.175***
	(0.011)	(0.112)	(0.006)	(0.010)	(0.021)
Mean of outcome	0.352	2.185	0.048	0.202	0.249
Observations	26,455	23,215	27,011	27,011	27,011
R-squared	0.128	0.143	0.047	0.034	0.032

* 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 province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2: Internet Effects by Education Levels



Notes: The omitted category is no schooling. 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.

Mechanism

- **Actual internet usage:**
  - social networking is the most important internet activity (44.5%), while only about 12% survey respondent uses internet for job search [1].
  - more internet access not only reduces the cost of acquiring job information online, but also communication cost among family and friends.

[1] Research ICT Africa (RIA), South Africa ICT access survey

- **Actual access variations:** accessing the internet for online job search is more difficult for the unskilled than the skilled.

Table 5: Impacts of Internet Connection on Cell and PC Ownership

	(1) Own a cellphone	(2) Own a computer	(3) Computer literate
% connected	0.096**	-0.005	-0.110**
	(0.044)	(0.009)	(0.048)
... × Beyond primary	-0.051	0.059***	0.147***
	(0.050)	(0.018)	(0.046)
Beyond primary	0.132***	-0.021**	0.029
	(0.026)	(0.008)	(0.019)
Observations	23,997	23,980	22,912
R-squared	0.575	0.600	0.704
Mean of outcome	0.686	0.044	0.257

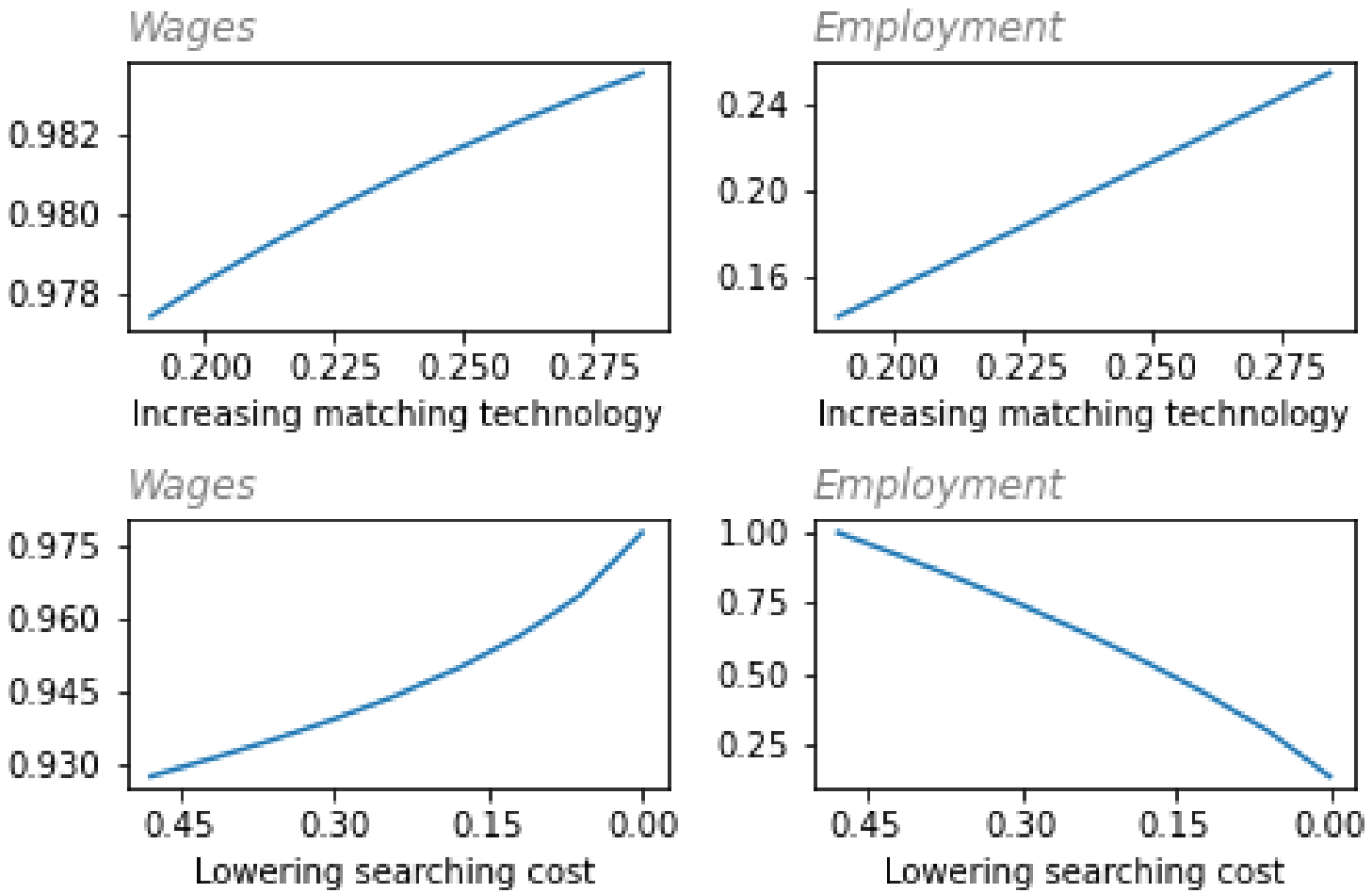
* Notes: Only workers between age 15 and 65 are included. All specifications include year, location fixed effects, age and gender control variables. Standard errors (in parentheses) are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Other considerations - general equilibrium impacts

Internet can affect both the labor supply and the labor demand, and the results on employment and wages should reflect the equilibrium outcomes.
Using the standard framework of unemployment and vacancies by Diamond, Mortensen and Pissarides (DMP):

- improvement in matching efficiency
- reduces the cost of learning about and applying for jobs

Figure 3: 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 Hagedoren et al.(2013).

Conclusion

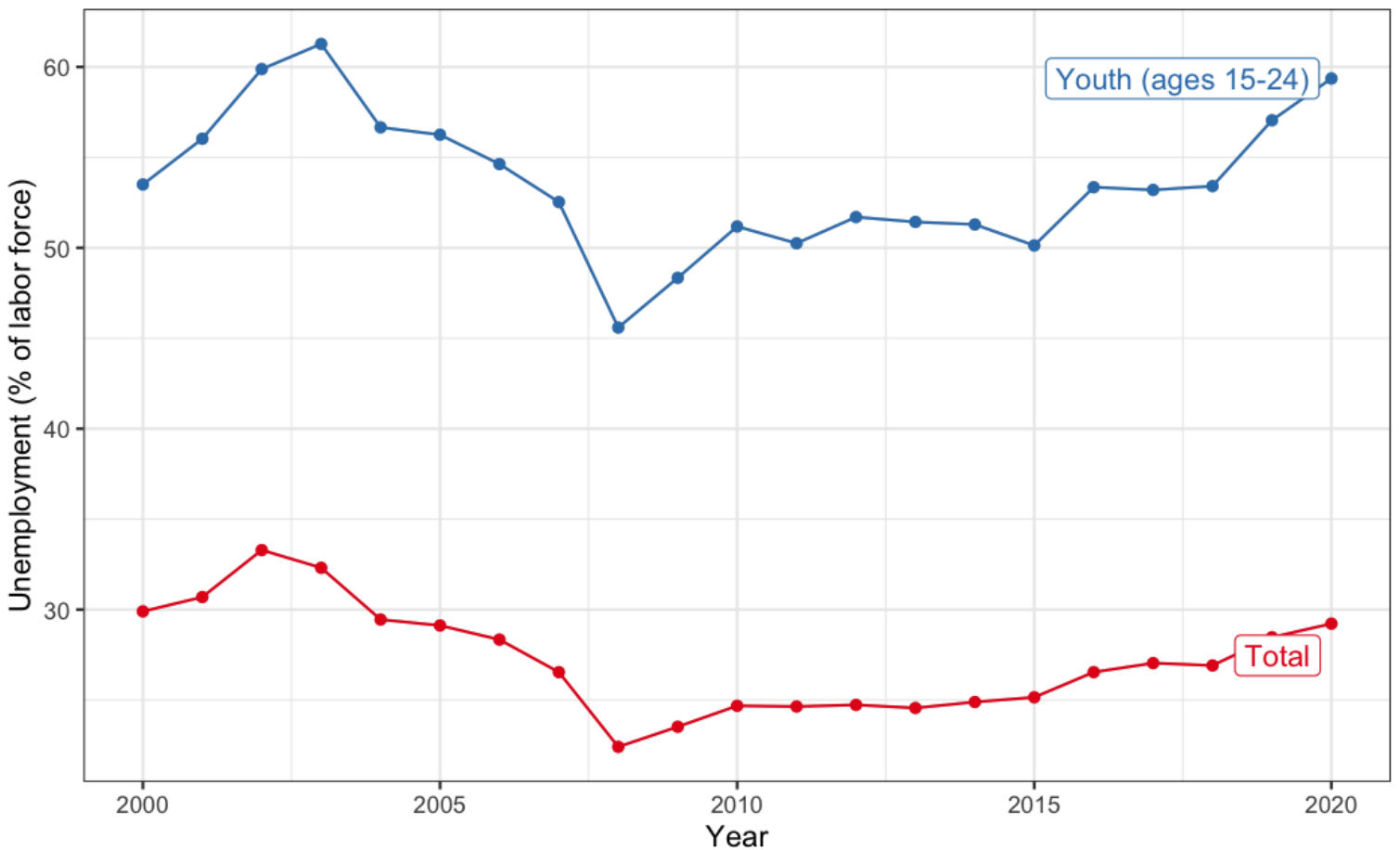
- Locations with better connectivity have higher employment rates and income, and the impact is driven by a significant increase in employment of skilled workers.
- With more available internet, all job seekers increase their use of personal network search, while only skilled ones increase the use of online search.

Policy implications:

- Complementary policies aiming at reducing costs, updating skill and digital literacy are critical for ensuring the overall benefits be shared broadly.

Go Back

Figure A1: Unemployment Rate in South Africa



Source: World Development Indicators.

Go Back

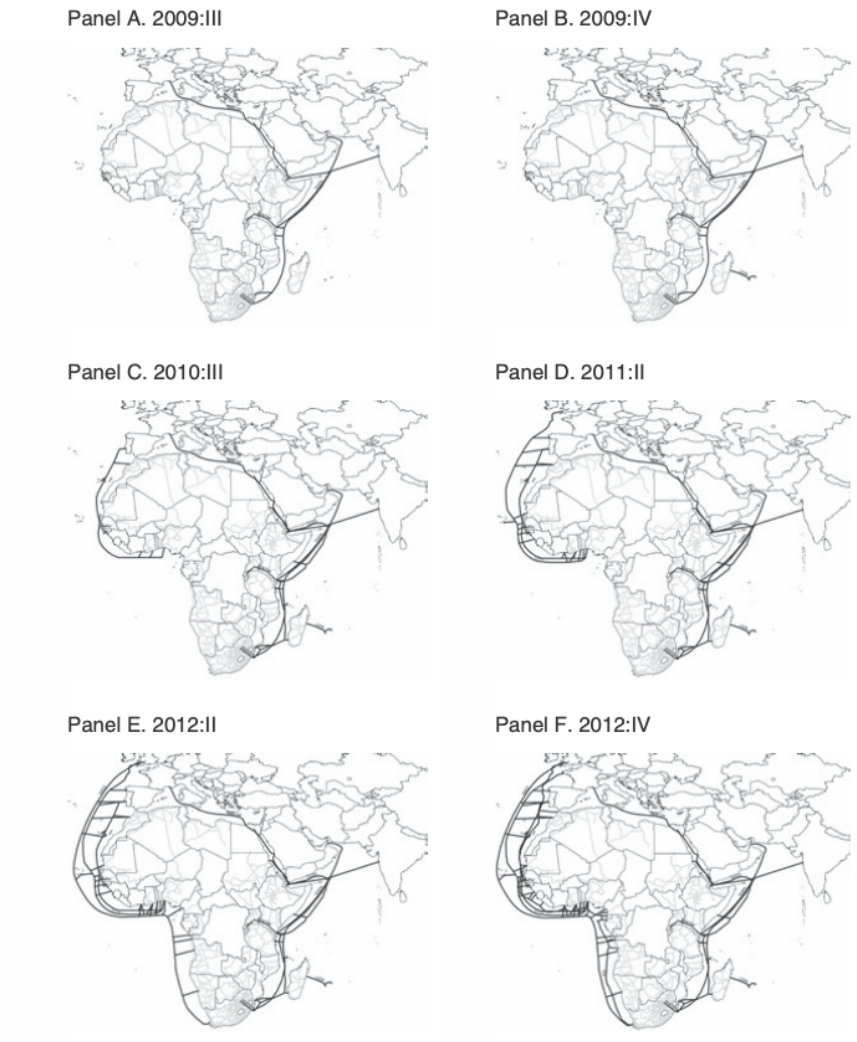
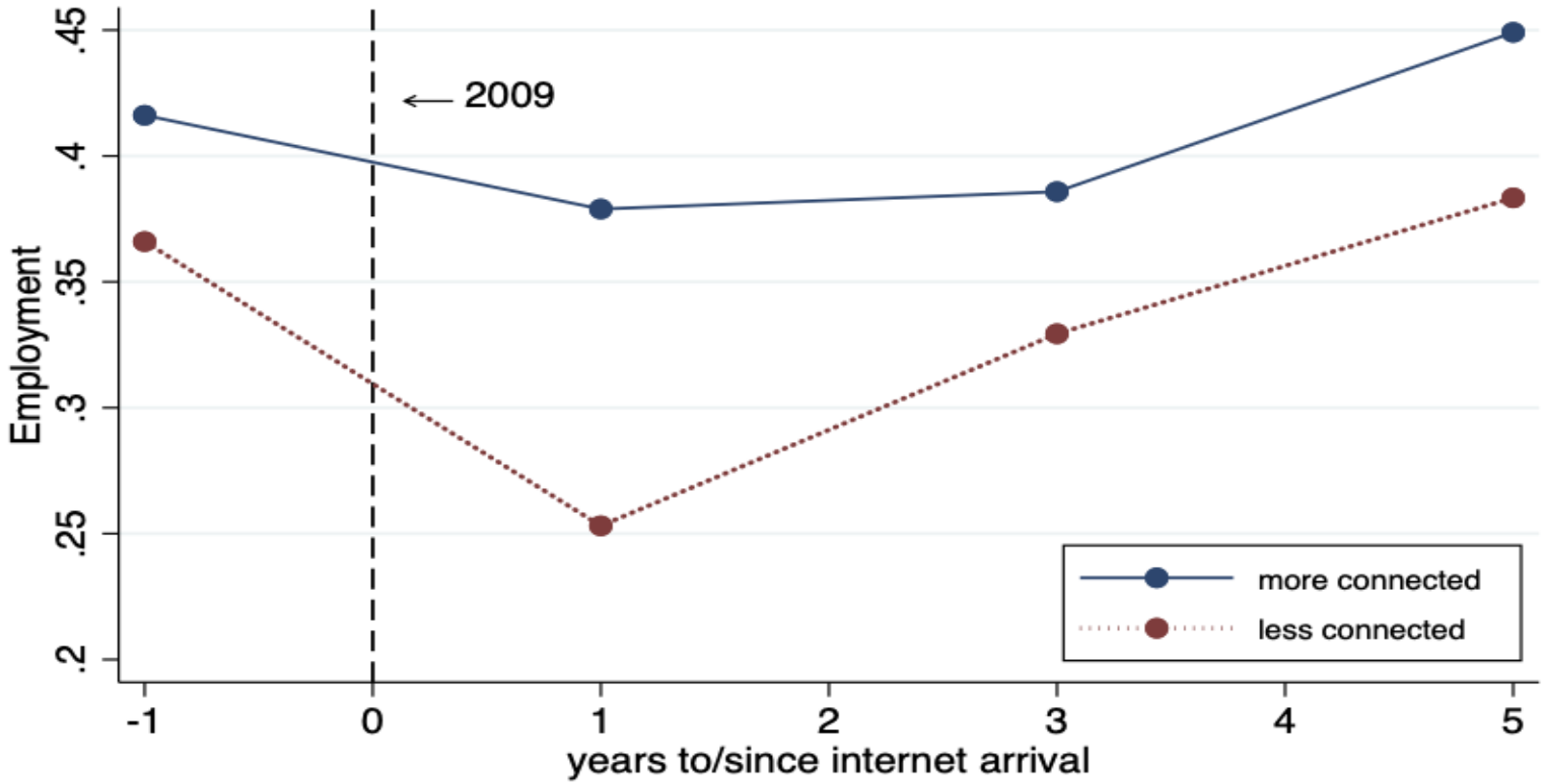


FIGURE 2. SUBMARINE INTERNET CABLE ARRIVAL IN AFRICA

[1] Source: Hjort & Poulsen (2019)

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



Notes: More (less) connected province is defined using the median internet penetration rate.

Figure A3: Internet speed by provinces

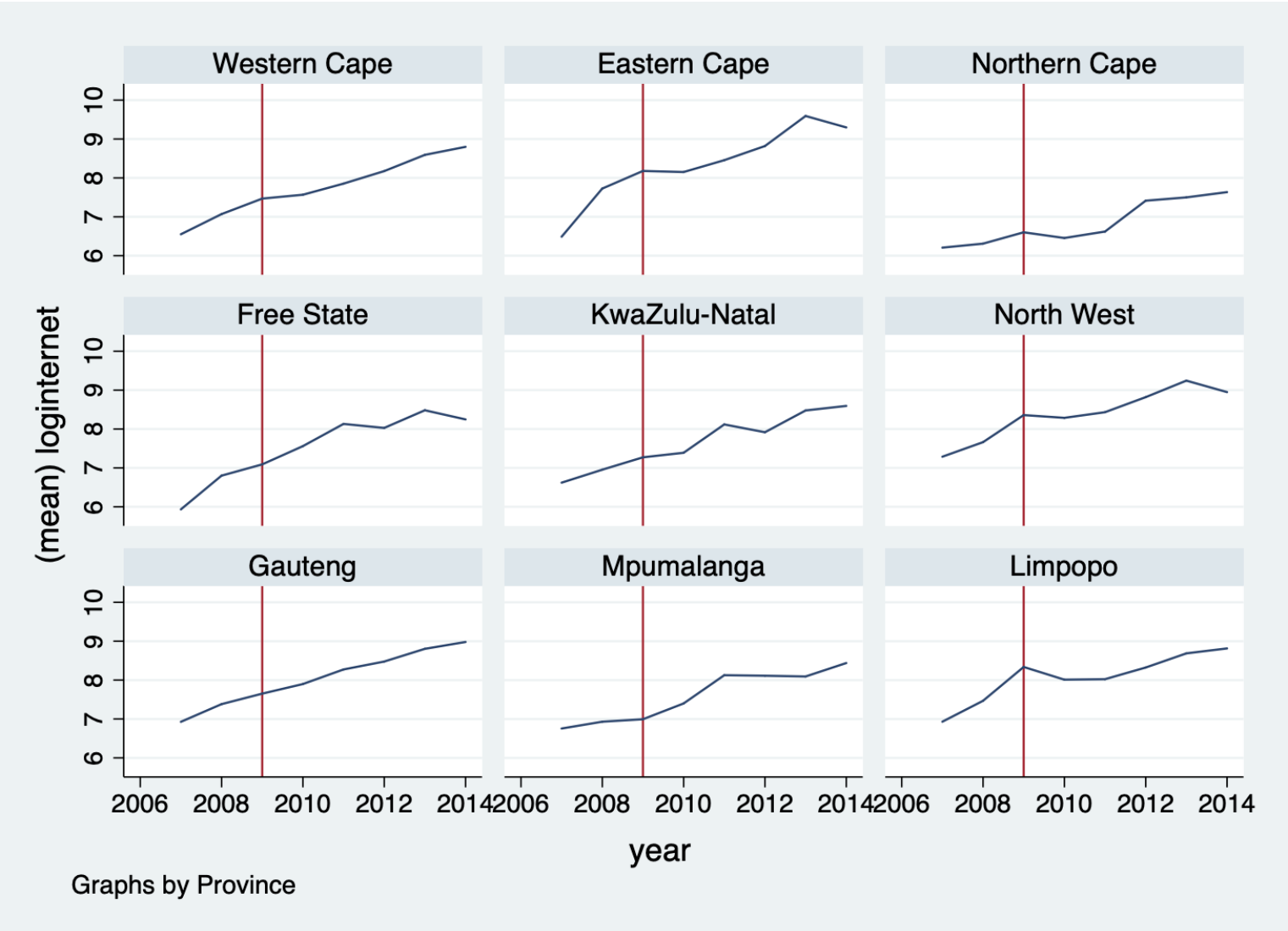


Table A1: Internet connection and lagged productivity variables

	(1) % connected	(2) % connected	(3) % connected
L.Employed	-0.004		
	(0.004)		
L.Education		-0.000	
		(0.001)	
L.Industry			-0.002
			(0.001)
Obs	16,296	8,954	3,920
R-squared	0.000	0.000	0.000

Notes: Dependent variable is the fast internet penetration rate. The small sample size for column 3 is because only a small portion of workers reported their industries. Robust standard errors (in parentheses). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Fixed Effects Estimates of Internet Connection on Job Outcomes

Outcome	(1) Employed (0/1)	(2) Total income (ihs)	(3) Salary wage (ihs)	(4) Weekly hours (log)	(5) Formal contract (0/1)
% connected	0.035	0.229	0.212	0.018	-0.001
	(0.031)	(0.202)	(0.345)	(0.218)	(0.061)
Mean of outcome	0.351	2.175	3.351	3.490	0.671
Observations	26,582	23,329	24,907	7,143	6,324
R-squared	0.100	0.105	0.084	0.057	0.032

* 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 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. Standard errors (in parentheses) are clustered at the province level. Control variables include age and gender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Heterogeneous effects

Table A3: Impacts of Internet Connection on Job Outcomes by Age

	(1) Employed	(2) Total income	(3) Online	(4) Network	(5) Num of methods
% connected	0.025	0.229	0.016	0.061**	-0.004
	(0.032)	(0.270)	(0.016)	(0.025)	(0.068)
.. × youth	0.038	0.057	-0.030***	0.007	0.114*
	(0.033)	(0.340)	(0.009)	(0.027)	(0.059)
Mean of outcome	0.351	2.175	0.047	0.194	0.240
Observations	26,582	23,329	27,998	27,998	27,998
R-squared	0.165	0.171	0.022	0.060	0.029

* 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 province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

DMP framework

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

$$H_t = A_t v_t^{\alpha} u_t^{1-\alpha} \tag{A1}$$

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 θ_t represents the labor market tightness.

All workers face the same constant unemployment risk λ. 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,

$$u = \frac{\lambda}{\lambda + A(\theta)^\alpha} \tag{A2}$$

Workers maximize the net present value of income and randomly search for vacant jobs while unemployed.

The flow value of being unemployed is $rU = b + A(\theta)^{\alpha}(W - U)$

The flow value of working is $rW = w + \lambda(U - W)$

Firms receive a flow value of profits for active jobs $rJ = p - w - \lambda J$

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,

$$p - w - \frac{(r + \lambda)c}{A(\theta)^{\alpha}} = 0 \tag{A3}$$

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.

$$w = \arg\max (W - U)^{\beta} (J - V)^{1-\beta}, \tag{A4}$$

where β 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,

$$w = (1 - \beta)b + \beta p(1 + c\theta) \tag{A5}$$

Equilibrium is a unique set of (u, θ, w) that satisfies the flow equilibrium condition A2, the job creation condition A3, and the wage equation A5.