Migration Costs, Selection, and Agricultural Productivity Gap

Qingen Gai [1] Naijia Guo [2] Bingjing Li [3] Qinghua Shi [4]
Xiaodong Zhu [5]

[1]Shanghai University of Finance and Economics
[2]Chinese University of Hong Kong
[3]National University of Singapore
[4]Shanghai Jiao Tong University
[5]University of Toronto

China Economic Summer Institute, August 2020

Agricultural Productivity Gap (APG)

- Cross-country differences in output per capita is mainly driven by the large differences in agriculture.
- There are large gaps in labor productivity between the agricultural and non-agricultural sectors in developing economies (Gollin, Parente, and Rogerson 2002, Caselli 2005, Restuccia, Yang, and Zhu 2008)
- The gaps remain to be large after controlling for observable differences in worker characteristics between the two sectors (Vollrath 2014, Gollin, Lagakos, and Waugh 2014)

Research Question

- What accounts for the large observable cross-sectional APG?
  - Unobservable differences in worker characteristics and sorting between the two sectors?
  - Barriers to worker mobility between the two sectors?
- We use unique data from China to investigate this question

APG in China

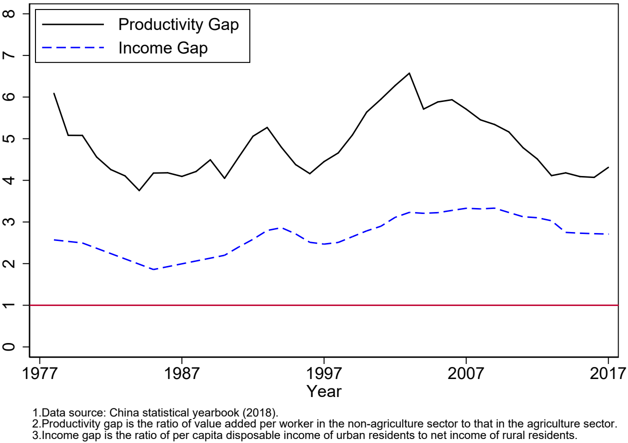


Figure: Raw Gap in China

Agricultural Productivity Gap: Conceptual Framework

$$Y_a = A_a H_a, Y_n = A_n H_n$$

- Agricultural productivity gap per efficiency unit of labor:

$$APG = \frac{P_n A_n}{P_a A_a}$$

This is not directly observable

- The cross-sectional agricultural productivity gap

$$CAPG = \frac{P_n Y_n / L_n}{P_a Y_a / L_a}$$

This is the one typically measured in the literature

- Decomposition

$$CAPG = APG \times \frac{H_n / L_n}{H_a / L_a}$$

$\frac{H_n/L_n}{H_a/L_a}$ is affected by worker selection.

Migration costs and Agricultural Productivity Gap

- If workers are homogeneous, then,
  - $CAPG = APG$
  - no-arbitrage condition implies

$$APG = \frac{P_n A_n}{P_a A_a} = \text{average migration costs}$$

- If workers are heterogeneous in unobserved abilities, $CAPG \neq APG$ due to selection
  - IV regression (LATE) estimates average migration cost of compilers
  - APG is the ATE, which cannot be directly estimated without imposing structural assumptions

What we do

We estimate APG and migration costs using panel data (National Fixed Point Survey) in rural China from 2003 to 2012:

- Use IV model to estimate migration costs non-parametrically
  - Use the roll-out of the New Rural Pension Scheme (NRPS) as an instrument
- Use a structural model to estimate (1) APG, (2) migration costs, and (3) selection bias
  - Agricultural and non-agricultural abilities
  - Differential returns to human capital between the two sectors
  - Heterogeneous migration costs
  - Use the NRPS as an IV

What we find

- Controlling for individual fixed effects does not have a big impact on the CAPG (from 48% to 52%) in China.
- Using the NRPS as an instrument, the estimated migration cost for compliers is 61% of non-agricultural earnings.
- Structural estimation
  - Average APG is 52%, larger for men and younger workers.
  - Migration cost accounts for 53% – 55% of annual non-agricultural earnings. Migration costs are larger for female, less educated, and older workers.
  - Comparative advantage is negatively correlated with the absolute advantage in the A sector.

Related Literature

- Agricultural productivity gap
  - FE model: Alvarez (2018), Hicks et al. (2018), Lagakos et al. (2020)
  - IV model: Lagakos et al (2020), Lagakos, Mobarak, and Waugh (2020)
  - Comparative advantage model: Lagakos and Waugh (2013), Young (2013), Pulido and Swiecki (2019)
- Misallocation and productivity
  - Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Adamopoulos and Restuccia (2014), Tombe and Zhu (2019), Bryan and Morten (2019)

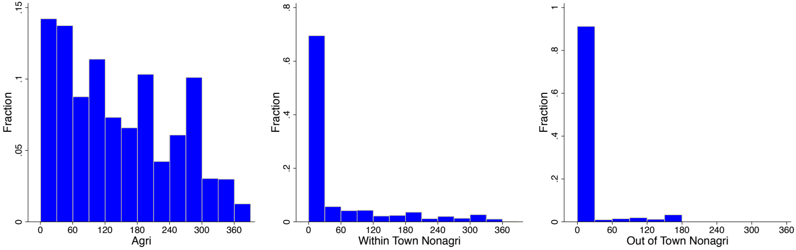
Data

National Fixed Point Survey (NFPS)

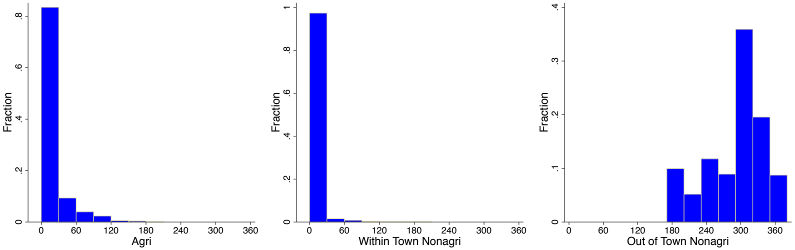
- NFPS is an annual panel survey collected by the Research Center of Rural Economy (RCRE) of the Chinese Ministry of Agriculture, beginning in 1986
- NFP villages were selected for representativeness based on region, income, cropping pattern, population and so on.
- The data set used in our analysis covers over 20,000 households from more than 350 villages in 31 provinces during 2003-2012.
  - NFP starts to collect individual information from 2003.
  - It surveys migrants and rural residents.
  - It collects data on household agriculture production, individual earnings and working days.

Key variables

- Individual agricultural earnings: allocate household's agricultural value-added to each member based on their annual agricultural working days.
  - Calculate the gross output for each type of crops as the production times the corresponding market price in that year
  - Value added is calculated using the sum of output of each type crops minus the sum of intermediate inputs of each type of crops.
- Individual non-agricultural earnings: days and annual earnings of working out of town
- Working in the non-agricultural sector (migration): working more than 180 days out of town.
- Earnings are deflated by prices from Brandt and Holz (2006).
- Restrict our sample to adults (20-54) without college degree

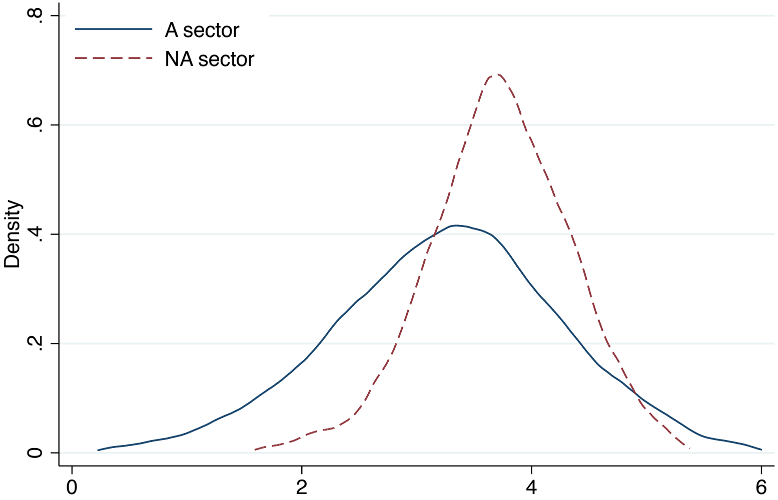


Panel A. Working days in different sectors for Agri workers



Panel B. Working days in different sectors for Nonagri workers

Distribution of daily earnings by sector



Agri sector: mean 3.26; sd 0.99

Non-Agri sector: mean 3.71; sd 0.61

Individual Fixed Effect Model

Empirical strategy

Inspired by Hendricks and Schoellman (2017), we focus on panel data with individual fixed effect to control for unobservable characteristics.

$$lnearning_{it} = \alpha Nonagri_{it} + \beta X_{it} + \delta_i + \lambda_t + \epsilon_{it}$$

- $lnearning_{it}$ is the log daily earning of individual i in year t
- $Nonagri_{it}$ is an indicator for individual i working in the non-agriculture sector in year t
- X_{it} is individual control variables
- δ_i, λ_t are individual and time fixed effects

APG

Table: APG: Pooled Cross-sectional OLS

Dep. Var.: ln Daily Earnings	(1)	(2)	(3)	(4)
NonAgri	0.4845***	0.4820***		
	(0.0148)	(0.0149)		
NonAgri w/in County			0.4494***	0.4484***
			(0.0161)	(0.0163)
NonAgri o/s County and w/in Province			0.4391***	0.4327***
			(0.0182)	(0.0182)
NonAgri o/s Province			0.5322***	0.5310***
			(0.0168)	(0.0169)
Individual controls	Y	Y	Y	Y
Province× Year FE	Y	Y	Y	Y
Village FE	Y	N	Y	N
Village× Year FE	N	Y	N	Y
Observations	208,246	208,231	208,246	208,231
R-squared	0.4540	0.5000	0.4546	0.5007

Notes: Robust standard errors are clustered at the village×year level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

APG

Table: APG: Individual Fixed Effects

Dep. Var.: ln Daily Earnings	(1)	(2)	(3)	(4)
NonAgri	0.5164***	0.5152***		
	(0.0170)	(0.0170)		
NonAgri w/in County			0.4994***	0.4980***
			(0.0179)	(0.0181)
NonAgri o/s County and w/in Province			0.4903***	0.4831***
			(0.0197)	(0.0197)
NonAgri o/s Province			0.5445***	0.5466***
			(0.0191)	(0.0191)
Individual controls	Y	Y	Y	Y
Individual FE	Y	Y	Y	Y
Province× Year FE	Y	N	Y	N
Village× Year FE	N	Y	N	Y
Observations	202,525	202,503	202,525	202,503
R-squared	0.7056	0.7410	0.7056	0.7411

Notes: Robust standard errors are clustered at the village×year level. *** p<0.01, ** p<0.05, * p<0.1

Comparison to the literature

Table: Agricultural productivity gap

	cross-section	individual-FE
China	0.48	0.52
Indonesia (Hicks et al. 2018)	0.70	0.25
Kenya (Hicks et al. 2018)	0.78	0.28
Brazil manufacturing/agricultural (Alvarez 2018)	0.48	0.09
Brazil service/agricultural (Alvarez 2018)	0.48	0.04
China (Lagakos et al 2020)	0.55	0.16
Ghana (Lagakos et al 2020)	0.41	0.15
Indonesia (Lagakos et al 2020)	0.63	0.15
Malawi (Lagakos et al 2020)	0.52	0.05
South Africa (Lagakos et al 2020)	0.74	0.21
Tanzania (Lagakos et al 2020)	0.67	0.11

Table: Evidence from CFPS

	Nominal consumption		Nominal consumption		Nominal consumption		Real income	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urban resident	0.5068***	0.0961**	0.3001***	0.1059**				
	(0.0117)	(0.0397)	(0.0147)	(0.0430)				
Non-agricultural sector					0.3433***	0.0529**	1.2767***	0.5270***
					(0.0108)	(0.0222)	(0.0203)	(0.0554)
Rural HuKou Only			Yes	Yes	Yes	Yes	Yes	Yes
Ind and HH Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes		Yes		Yes
Observations	88496	74061	63556	52227	41157	30469	21023	8557
R-squared	0.0810	0.6822	0.0256	0.6344	0.0384	0.6492	0.2128	0.7359

Standard errors in parentheses

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

Selection bias

- Individual FE regression does not control for all selection biases
  - Workers that switch out of agriculture are the ones who had smaller migration costs and/or larger comparative advantage
  - Change in migration costs could be endogenous
  - If migration costs decline more for individuals with larger comparative advantage, the FE estimate would have an upward bias.

Instrumental Variable Model

- Instrument: New Rural Pension Scheme (NRPS)

IV: NRPS

The roll-out of the NRPS

- First round of pilot counties began the NRPS in 2009, the second round in 2010, and the rest two in 2011 and 2012.
- Starting from the pension eligible age of 60, the pension benefits for a pensioner are the sum of the accumulated total funds in the individual account, plus the basic pension benefits (55 yuan per month).
- For adults aged below 60, the NRPS only affects the migration costs, but not incomes in the two sectors directly.

NRPS

- How would the NRPS affect the migration costs?
  - Make the elderly more independent and less likely to rely on the eldercare provided by children (Zhang et al., 2017, Guo et al., 2019), increase the healthcare service consumption for the elderly and market care (Zhang and Chen, 2014, Eggleston et al., 2016, Chen et al., 2017).
  - Reduce the labor supply and the time in farm work for the elderly, and therefore increase the time they can spend on their grandchildren (Jiao 2016 and Huang and Zhang, 2016).
  - Affect the credit constraint of potential migrants (Zhang et al., 2013, Huang and Zhang, 2016).

Empirical Evidence

IV Model

- First stage:

$$\begin{aligned} NonAgri_{ihjt} = & \beta_1 Elder60_{hjt} \times NPRS_{jt} + \beta_2 NPRS_{jt} + \beta_3 Elder60_{hjt} \\ & + X'_{ihjt}\beta_4 + D_h + D_{pt} + \varepsilon_{ihjt} \end{aligned}$$

- Second stage:

$$\begin{aligned} \ln y_{ihjt} = & \gamma_1 \widehat{NonAgri}_{ihjt} + \gamma_2 NPRS_{jt} + \gamma_3 Elder60_{hjt} \\ & + X'_{ihjt}\gamma_4 + D_h + D_{pt} + \nu_{ihjt} \end{aligned}$$

- Excluded IV: $Elder60_{hjt} \times NPRS_{jt}$ (Triple Differences)
- Identification assumption: $Elder60_{hjt} \times NPRS_{jt}$ is uncorrelated with ν_{ihjt}.

IV results

Table: Sector of Employment and Daily Earnings: IV Approach

	(1)	(2)	(3)	(4)
Dep. Var.:	NonAgri	ln Daily Earnings	ln Daily Earnings	ln Daily Earnings
	First Stage	Reduced Form	2SLS	OLS
Nonagri			0.6106*	0.4845***
			(0.3567)	(0.0148)
Elder60×NPRS	0.0427***	0.0261*		0.0054
	(0.0087)	(0.0157)		(0.0152)
NRPS	0.0041	-0.0061	-0.0086	-0.0081
	(0.0083)	(0.0294)	(0.0298)	(0.0294)
Elder60	0.0218***	0.0234***	0.0101	0.0128**
	(0.0027)	(0.0056)	(0.0113)	(0.0055)
Individual and household controls	Y	Y	Y	Y
Province× Year FE	Y	Y	Y	Y
Village FE	Y	Y	Y	Y
Observations	208,246	208,246	208,246	208,246
R-squared	0.3548	0.4149	0.0930	0.4540
Kleibergen-Paap F-Stat	–	–	24.32	–

Notes: Individual controls include all independent variables in Table **??**. Robust standard errors are clustered at the village×year level. *** p<0.01, ** p<0.05, * p<0.1

Simple Comparative Advantage Model

Model setup

- Each worker is endowed with a vector of individual productivities (z_A, z_N), following joint normal distribution:

$$\begin{pmatrix} z_A \\ z_N \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & \sigma_{AN} \\ \sigma_{AN} & \sigma_N^2 \end{pmatrix}\right)$$

- R is APG and m is migration cost
- Log income net of migration cost in the NA sector is $\ln m + \ln R + z_N = r - \mu + z_N$, where $r = \ln R$ and $\mu = -\ln m$
- Log income in the A sector is z_A
- Denote $v = z_N - z_A$, and $\sigma_v^2 = \sigma_N^2 - 2\sigma_{NA} + \sigma_A^2$.

- The share of workers in the NA sector is given by:

$$\pi_N = \Pr\left[z_N - z_A > \mu - r\right] = \Phi\left(\frac{r-\mu}{\sigma_v}\right).$$

- Observed average income of workers in the A sector

$$E(z_A | \underbrace{z_N - z_A < \mu - r}_{\text{workers in the A sector}}) = -\sigma_A \rho_{Av} \frac{\phi(\frac{r-\mu}{\sigma_v})}{1 - \Phi(\frac{r-\mu}{\sigma_v})}$$

- Observed average income of workers in the NA sector

$$E(r + z_N | \underbrace{z_N - z_A > \mu - r}_{\text{workers in the NA sector}}) = r + \sigma_N \rho_{Nv} \frac{\phi(\frac{r-\mu}{\sigma_v})}{\Phi(\frac{r-\mu}{\sigma_v})}$$

- Observed cross-sectional APG (CAPG)

$$E(r+z_N|\underbrace{z_N - z_A > \mu - r}_{\text{workers in the NA sector}}) - E(z_A|\underbrace{z_N - z_A < \mu - r}_{\text{workers in the A sector}})$$

$$= r + \sigma_N \rho_{Nv} \frac{\phi(\frac{r-\mu}{\sigma_v})}{\Phi(\frac{r-\mu}{\sigma_v})} + \sigma_A \rho_{Av} \frac{\phi(\frac{r-\mu}{\sigma_v})}{1-\Phi(\frac{r-\mu}{\sigma_v})}.$$

- We want to recover APG (ATE), which is r

$$ATE = E[r + z_N - z_A] = r$$

- Sorting can result in a biased estimate of r

- IV estimate is LATE

$$
\begin{aligned}
&E(r + z_N - z_A | \mu' < r + z_N - z_A < \mu) \\
&= r + \sigma_v E(\frac{z_N - z_A}{\sigma_v} | \frac{\mu' - r}{\sigma_v} < \frac{z_N - z_A}{\sigma_v} < \frac{\mu - r}{\sigma_v}) \\
&= r + \sigma_v \frac{\phi(\frac{r-\mu'}{\sigma_v}) - \phi(\frac{r-\mu}{\sigma_v})}{\Phi(\frac{r-\mu'}{\sigma_v}) - \Phi(\frac{r-\mu}{\sigma_v})}.
\end{aligned}
$$

Denote $\Delta = \frac{\mu - \mu'}{\sigma_v}$. In an extreme case that changes in migration costs are infinitely small ($\Delta \to 0$), the LATE estimate can be rewritten as

$$
\begin{aligned}
&\lim_{\Delta \to 0} r + \sigma_v \frac{\phi(\frac{r-\mu}{\sigma_v} + \Delta) - \phi(\frac{r-\mu}{\sigma_v})}{\Phi(\frac{r-\mu}{\sigma_v} + \Delta) - \Phi(\frac{r-\mu}{\sigma_v})} \\
&= r + \sigma_v \frac{\phi'(\frac{r-\mu}{\sigma_v})}{\phi(\frac{r-\mu}{\sigma_v})} \\
&= r + \sigma_v \frac{d \ln \phi(\frac{r-\mu}{\sigma_v})}{d(\frac{r-\mu}{\sigma_v})} \\
&= \mu.
\end{aligned}
$$

We prove that this property holds for general distribution of (z_A, z_N).

- When $\Delta \to 0$, the LATE estimate recovers the migration cost μ.
- Intuitively, marginal workers have $r + z_N - z_A = \mu$.
- If $\Delta > 0$, we get a lower bound estimate of migration costs.

The difference between OLS and IV estimates is

$$E(r + z_N | \underbrace{z_N - z_A > \mu - r}_{\text{workers in the NA sector}}) - E(z_A | \underbrace{z_N - z_A < \mu - r}_{\text{workers in the A sector}})$$

$$= r + \mu + \sigma_N \rho_{Nv} \frac{\phi(\frac{r-\mu}{\sigma_v})}{\Phi(\frac{r-\mu}{\sigma_v})} + \sigma_A \rho_{Av} \frac{\phi(\frac{r-\mu}{\sigma_v})}{1 - \Phi(\frac{r-\mu}{\sigma_v})}.$$

Sufficient conditions for OLS $<$ IV are

- $\sigma_A > \sigma_N$
- migration costs sufficiently large
- Abramitzky et al 2012 and Lagakos et al 2020

Extended Comparative Advantage Model

- Differential returns to human capital between the two sectors
- Heterogeneous migration costs
- Use the NRPS as an IV

Model setup

- The annual income an individual receives in sector s at time t is:

$$y_t^s = R_t^s H_{it}^s h^s$$

where R_t^s is the price human capital (daily). h^s is the annual working days and we set $h^A = 208$ and $h^N = 292$.

- Human capital of individual i in sector s at time t, defined as

$$\log H_{it}^s = X_{it}'\beta^s + z_i^s + \epsilon_{it}^s$$

where β^s are the differential returns to human capital by sector. X_{it} includes gender, years of schooling, age, and age square. z_i^s is the ability of worker i in sector s. ϵ_{it}^s are i.i.d shocks.

Model setup

- The value of working in sector s is

$$V^s = \ln y_t^s - \phi_{it}^N 1\{s_{i,t} = N\} - \phi_{it}^A 1\{s_{i,t} = N, s_{i,t-1} = A\}$$

  - ϕ_{it}^N is an annual migration cost, defined as

    $$\phi_{it}^N = Z_{it}' \gamma^N + \zeta_{it}^N$$

    where Z_{it} includes gender, education, age, age square, NRPS, elderly, NRPS*elderly. ζ_{it}^N is an i.i.d. shock that follows the standard normal distribution.
  - ϕ_{it}^A is a one-time migration cost, defined as

    $$\phi_{it}^A = Z_{it}' \gamma^A + \zeta_{it}^A$$

    where ζ_{it}^A is an i.i.d. shock that follows the standard normal distribution.

Migration costs

What contributes to the migration costs?

- Annual migration costs
  - Utility loss because of longer hours
  - Hukou: higher costs of education and health care
  - Physical separation from family
  - Amenities in the city
- One-time migration costs
  - Temporary unemployment caused by labor market friction
  - Travel costs from the home village to the destination city

Individual migration choice

- Workers choose the sector $s \in \{A, N\}$ at each time t to maximize their contemporaneous utility. The worker's problem is:

$$V(\Omega_{it}) = \max_s \{EV^s\}$$

Individuals observe the shocks on migration costs, but not income shocks when they make the migration decision.

  - Out migration and return migration decisions

$$OM_{it} = \begin{cases} 1 & \text{if } \ln E[y_{it}^N] - \ln E[y_{it}^A] - \phi_{it}^A - \phi_{it}^N \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$RM_{it} = \begin{cases} 1 & \text{if } \ln E[y_{it}^A] - \ln E[y_{it}^N] - \phi_{it}^N \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Identification strategy: NRPS

- We obtain the identification of our Roy model by using the panel data and an instrumental variable.
  - Heckman and Honore (1990) show that the estimation of selection models using only cross-sectional data faces identification challenges.
  - Pulido and Swiecki (2019) prove that with longitudinal data and functional form assumptions, the Roy model is identified.
  - An instrument that only affects the migration costs, but not earnings helps to better identify the migration costs without relying on the functional form assumptions.

Estimation

We use Maximum Likelihood to estimate the model

$$L = \prod_i \prod_t \int \phi_{\epsilon^s}(\ln y_{it}^s - \ln(R_{it}^s d^s) - X_{it}'\beta^s - \theta_i^s)$$
$$\Phi_{\xi^A}(X_{it}'(\beta^N - \beta^A) + \theta_i^N - \theta_i^A - Z_{it}'(\gamma^N + \gamma^A))^{1\{OM_{it}=1\}}$$
$$(1 - \Phi_{\xi^A}(X_{it}'(\beta^N - \beta^A) + \theta_i^N - \theta_i^A - Z_{it}'(\gamma^N + \gamma^A)))^{1\{OM_{it}=0\}}$$
$$\Phi_{\xi^N}(X_{it}'(\beta^A - \beta^N) + \theta_i^A - \theta_i^N - Z_{it}'\gamma^N)^{1\{RM_{it}=1\}}$$
$$(1 - \Phi_{\xi^N}(X_{it}'(\beta^A - \beta^N) + \theta_i^A - \theta_i^N - Z_{it}'\gamma^N))^{1\{RM_{it}=0\}} d\Phi_{\theta_i^A, \theta_i^N}$$

where ϕ_x and Φ_x are the pdf and cdf of Normal distribution for x.
$\xi_{it}^A = \zeta_{it}^A - \zeta_{it}^N + \epsilon_{it}^A - \epsilon_{it}^N$. $\xi_{it}^N = \zeta_{it}^N + \epsilon_{it}^N - \epsilon_{it}^A$

Structural estimation results

	Agricultural	Non-agricultural	Nonagri - Agri
Log price for human capital	2.303	2.822	0.519
	(4.21)	(3.78)	
Human capital – year trend	0.091	0.091	0.000
	(4.22)	(5.17)	
Human capital – male	0.152	0.269	0.116
	(3.42)	(3.66)	
Human capital – years of education	0.013	0.033	0.020
	(10.98)	(12.42)	
Human capital – age	0.117	0.075	-0.042
	(6.34)	(8.52)	
Human capital – age square	(0.001)	(0.001)	0.000
	(2.91)	(2.34)	
Human capital – access to migration	(1.352)	0.607	1.959
	(10.29)	(11.72)	
Sd of wage shock	0.740	0.363	
	(6.27)	(3.56)	
Sd of ability	0.699	0.453	
	(4.99)	(4.25)	
Correlation between abilities	0.466		
	(3.19)		

t-statistics in parentheses.

Structural estimation results

- APG is 51.9%, and larger for men and younger workers.
- Agri and non-agri abilities are positively correlated ($\rho = 0.47$).
- Agricultural ability has a larger variation than non-agricultural ability (0.70 vs. 0.45).

Structural estimation results

	Annual migration cost	One-time migration cost	Out migration cost
Constant	-0.534	-0.017	-0.551
	(6.66)	(9.45)	
Male	-0.179	-0.280	-0.458
	(6.13)	(5.88)	
Years of education	-0.053	0.024	-0.029
	(11.91)	(7.75)	
Age	0.059	0.086	0.145
	(8.25)	(6.29)	
Age square	-0.001	-0.001	-0.002
	(11.29)	(6.52)	
Access to migration	-2.854	0.255	-2.600
	(11.99)	(6.65)	
NRPS	-0.091	0.092	0.001
	(4.00)	(3.75)	
Elderly above 60	-0.052	0.047	-0.005
	(4.38)	(5.21)	
NRPS * elderly	-0.102	-0.053	-0.155
	(6.22)	(5.45)	

t-statistics in parentheses.

Out migration cost is equal to the one-time migration cost plus the annual migration cost.

Structural estimation results

- On average, annual migration costs account for 53% of annual earnings and out migration costs account for 55%.
- The annual and out migration costs are larger for female, less educated, and older workers.
- The NRPS reduces annual and one-time migration costs, especially for those with an elderly above 60.

Positive selection or negative selection?

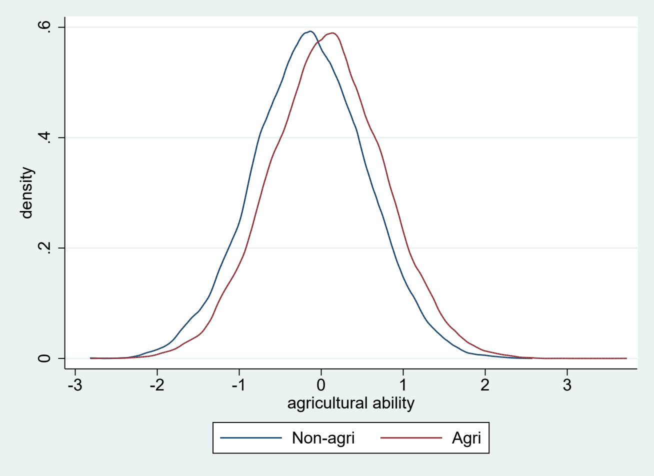


Figure: Agricultural ability of workers in the agri and non-agri sectors

Positive selection or negative selection?

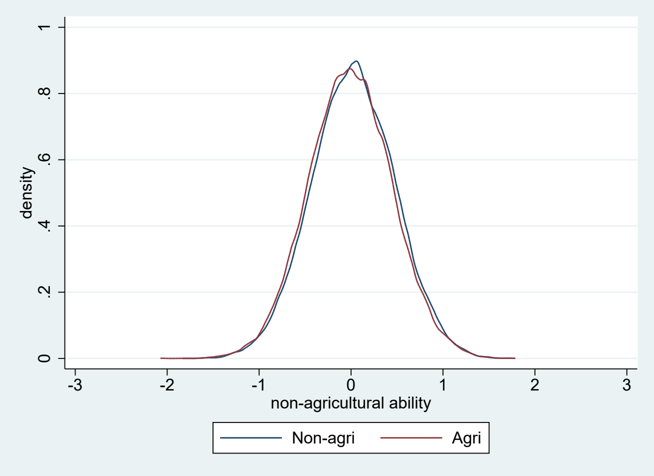


Figure: Non-agricultural ability of workers in the agri and non-agri sectors

Positive selection or negative selection?

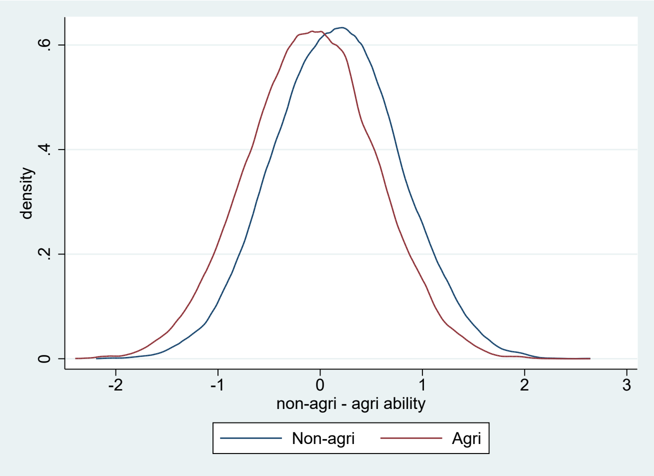


Figure: Relative ability (non-agri minus agri) of workers in the agri and non-agri sectors

Positive selection or negative selection?

- Selection in terms of agricultural ability, but not much in terms of non-agricultural ability
- Negative selection: comparative advantage is negatively correlated with the absolute advantage in the A sector
  - Based on our estimates, $Cov(z_N - z_A, z_N) > 0$ and $Cov(z_N - z_A, z_A) < 0$.
  - Hence $\sigma_A > \sigma_N$, implying IV > OLS.

Conclusion

Using a nationally representative long-term panel data (NFP) from China, this paper finds:

- Reduced form estimates: 48% – 61%.
  - IV estimate recovers migration costs, not APG
  - OLS and FE could be biased downward or upward, depending on the distribution of abilities and migration costs
- Structural model
  - APG is 52%
  - Migration cost accounts for 53% – 55% of annual non-agricultural earnings
  - Negative selection
- Future plan
  - General equilibrium model to analyze the impact of reducing migration costs

Thanks!

Table: Summary Statistics

Sample:	All	Non-agri	Agri
ln Daily wage	3.401	3.717	3.264
	(0.920)	(0.618)	(0.993
ln Annual income	8.680	9.373	8.380
	(1.064)	(0.630)	(1.074)
Total working days	236.882	303.588	208.051
	(101.624)	(44.114)	(105.809)
Share of working days in:			
Within-town agri production	0.556	0.036	0.781
	(0.435)	(0.078)	(0.318)
Within-town non-agri production	0.122	0.006	0.173
	(0.258)	(0.032)	(0.294)
Out-of-town	0.322	0.958	0.046
	(0.443)	(0.086)	(0.162)
Age	37.987	31.849	40.640
	(10.091)	(8.861)	(9.412)
Years of Schooling	7.273	8.152	6.893
	(2.431)	(2.046)	(2.485)
Female	0.470	0.330	0.531
	(0.499)	(0.470)	(0.499)
Poor health status	0.012	0.003	0.015
	(0.107)	(0.057)	(0.122)
Agricultural Hukou	0.977	0.962	0.983
	(0.151)	(0.192)	(0.129)
Arable land per capita	2.195	1.401	2.538
	(2.870)	(1.695)	(3.188)
Household with an elderly aged $\geq$60	0.278	0.343	0.249
	(0.448)	(0.475)	(0.433)
Number of observations	238312	71918	166394

Notes: Standard deviation in parentheses.

Sample of NFP

Figure: Village's distribution of NFP

NRPS

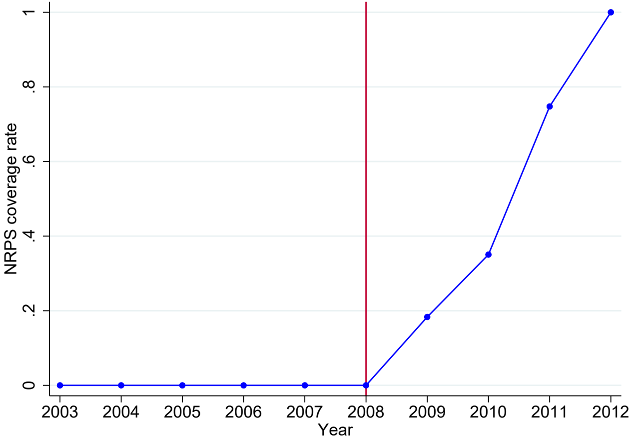


Figure: NRPS coverage rate

Table: Distribution of the Number of Years Individuals Can be Tracked

	Raw Data		Our Sample	
Years	Observations	Share(%)	Observations	Share(%)
1	51,414	30.46		
2	24,416	14.47	13,671	26.08
3	16,507	9.78	9,453	18.03
4	13,707	8.12	7,234	13.80
5	10,087	5.98	5,463	10.42
6	8,604	5.10	4,557	8.69
7	7,458	4.42	3,644	6.95
8	8,766	5.19	3,396	6.48
9	11,867	7.03	3,061	5.84
10	15,946	9.45	1,936	3.69
Total	168,772	100	52,415	100

Table: Summary Statistics: Labor Allocation and Sector of Employment by Out-of-town Labor Supply

	Agri Sector		Non-Agri Sector
Sample:	0 day	$(0, 180]$ days	> 180 days
Total working days	205.318	234.925	303.587
	(107.939)	(76.977)	(44.113)
Share of working days in:			
Within-town agri production	0.817	0.427	0.036
	(0.303	(0.228)	(0.078)
Within-town non-agri production	0.183	0.071	0.006
	(0.303	(0.150)	(0.032)
Out-of-town	0.000	0.502	0.958
	(0.000	(0.236)	(0.086)
(Self-reported) Non-agricultural sector	0.185	0.387	0.918
	(0.388	(0.487)	(0.274)
ln Daily wage in Non-agricultural sector	0.000	3.531	3.463
	(0.000	(0.682)	(0.662)
ln Daily wage in agricultural sector	2.998	2.894	2.994
	(1.009	(0.997)	(1.032)
Number of observations	151,031	15,364	71,917

Notes: Standard deviation in parentheses.

Table: Observable Characteristics and Migration Status

Dep. Var.:	(1) NonAgri All	(2) NonAgri All	(3) NonAgri w/in County	(4) NonAgri o/s County w/in Province	(5) NonAgri o/s Province
Age $\in [30, 40)$	-0.1970***	-0.1970***	-0.0127***	-0.0678***	-0.1165***
	(0.0041)	(0.0041)	(0.0023)	(0.0029)	(0.0037)
Age $\in [40, 50)$	-0.3582***	-0.3571***	-0.0402***	-0.1164***	-0.2015***
	(0.0050)	(0.0050)	(0.0025)	(0.0038)	(0.0049)
Age $\in [50, 55)$	-0.4291***	-0.4265***	-0.0632***	-0.1331***	-0.2326***
	(0.0057)	(0.0057)	(0.0028)	(0.0039)	(0.0056)
Primary School	0.0489***	0.0502***	0.0078	-0.0124***	0.0534***
	(0.0082)	(0.0083)	(0.0048)	(0.0045)	(0.0085)
Middle School	0.1015***	0.1027***	0.0173***	0.0020	0.0822***
	(0.0083)	(0.0084)	(0.0049)	(0.0047)	(0.0088)
High School	0.1429***	0.1430***	0.0306***	0.0308***	0.0816***
	(0.0088)	(0.0089)	(0.0052)	(0.0052)	(0.0090)
Female	-0.1363***	-0.1350***	-0.0393***	-0.0427***	-0.0542***
	(0.0029)	(0.0029)	(0.0017)	(0.0016)	(0.0020)
Poor Health	-0.0834***	-0.0839***	-0.0159***	-0.0186***	-0.0489***
	(0.0075)	(0.0075)	(0.0036)	(0.0034)	(0.0059)
Area of Arable Land	-0.0089***	-0.0086***	-0.0033***	-0.0021***	-0.0035***
	(0.0008)	(0.0008)	(0.0004)	(0.0005)	(0.0005)
Urban Hukou	0.0719***	0.0698***	0.0460***	0.0493***	-0.0234***
	(0.0077)	(0.0078)	(0.0060)	(0.0063)	(0.0066)
Elder60	0.0282***	0.0278***	0.0014	0.0109***	0.0159***
	(0.0026)	(0.0026)	(0.0016)	(0.0018)	(0.0022)
Lagged Access	0.6142***	–	0.2943***	0.1618***	0.1580***
	(0.0393)	–	(0.0303)	(0.0263)	(0.0297)
Province× Year FE	Y	N	Y	Y	Y
Village FE	Y	N	Y	Y	Y
Village× Year FE	N	Y	N	N	N
Observations	208,246	208,231	208,246	208,246	208,246
R-squared	0.3545	0.3698	0.1597	0.1663	0.2819

Notes: Robust standard errors are clustered at the village×year level. *** p<0.01, ** p<0.05, * p<0.1

APG: Annual Earnings

Table: APG: Annual Income

Dep. Var.: ln Annual Income	(1) OLS	(2) OLS-FE	(3) Reduced Form	(4) 2SLS
NonAgri	1.0496***	1.1192***		1.0591***
	(0.0137)	(0.0168)		(0.3570)
*Elder*60 × *NPRS*			0.0452***	
			(0.0166)	
NPRS	-0.0446	-0.0701**	-0.0404	-0.0448
	(0.0277)	(0.0281)	(0.0291)	(0.0287)
*Elder*60	0.0231***	0.0449***	0.0460***	0.0228*
	(0.0055)	(0.0109)	(0.0067)	(0.0119)
Individual controls	Y	Y	Y	Y
Province× Year FE	Y	Y	Y	Y
Village FE	Y	N	Y	Y
Individual FE	N	Y	N	N
Observations	208,246	208,247	208,246	208,246
R-squared	0.4508	0.7267	0.3155	0.2612
Kleibergen-Paap F-Stat				24.32

Notes: Individual controls include all independent variables in Table **??**. Robust standard errors are clustered at the village×year level. *** p<0.01, ** p<0.05, * p<0.1

Mechanisms: How NRPS Affect Migration?

Table: NPRS and the Elderly Labor Supply

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	All	Age$<$70	All	Age$<$70	All	Age$<$70
Dep. Var.:	Working days	Working days	Working days$>$ 120	Working days$>$ 120	Annual income	Annual income
	Poisson	Poisson	OLS	OLS	Poisson	Poisson
NPRS	-0.0063	-0.0380	-0.0123	-0.0318**	-0.0446	-0.0740*
	(0.0286)	(0.0279)	(0.0130)	(0.0146)	(0.0379)	(0.0391)
Individual controls	Y	Y	Y	Y	Y	Y
Province$\times$ Year FE	Y	Y	Y	Y	Y	Y
Village FE	Y	Y	Y	Y	Y	Y
Observations	58,785	45,825	58,813	45,835	58,756	45,803
R-squared	–	–	0.3151	0.2976	–	–

Notes: All columns restrict the sample to the elderly with medium or better health status (i.e. $HealthStatus \leq 3$). Individual level controls include dummies for age groups (60-64, 65-69, 70-74, and $>$ 75), dummies for educational attainment (illiterate, primary school, middle school, high school and college), gender, dummies for health status, arable land per capita, and type of Hukou. Robust standard errors are clustered at the village$\times$year level. *** p$<$0.01, ** p$<$0.05, * p$<$0.1, † p$<$0.15

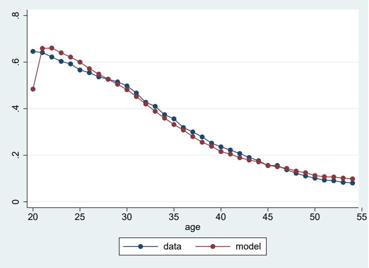
Mechanisms: How NRPS Affect Migration?

Table: NRPS and Medical Expenditure

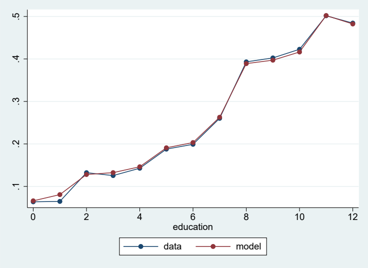
Dep. Var.: ln (1+Medical Expenditure)	(1) OLS	(2) OLS	(3) OLS
NRPS	0.3349**		
	(0.1464)		
$\mathbf{1}$(HealthStatus$\leq$ 3)$\times$ NRPS		0.3158**	
		(0.1487)	
$\mathbf{1}$(HealthStatus$\geq$ 4)$\times$ NRPS		0.4072**	
		(0.1688)	
$\mathbf{1}$(HealthStatus$=$ 1)$\times$ NRPS			0.1488
			(0.1768)
$\mathbf{1}$(HealthStatus$=$ 2)$\times$ NRPS			0.3202**
			(0.1599)
$\mathbf{1}$(HealthStatus$=$ 3)$\times$ NRPS			0.4755***
			(0.1655)
$\mathbf{1}$(HealthStatus$=$ 4)$\times$ NRPS			0.3943**
			(0.1786)
$\mathbf{1}$(HealthStatus$=$ 5)$\times$ NRPS			0.4473**
			(0.2006)
Individual and household controls	Y	Y	Y
Province$\times$ Year FE	Y	Y	Y
Village FE	Y	Y	Y
Observations	74,951	74,951	74,951
R-squared	0.2602	0.2603	0.2604

Back

Model fit



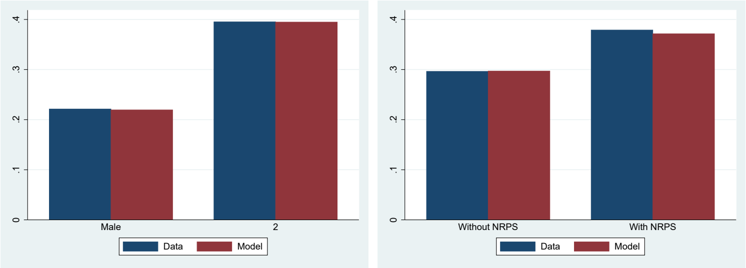
(a) By age



(b) By education

Figure: Fraction in the non-agricultural sector

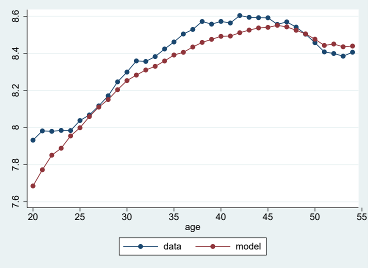
Model fit



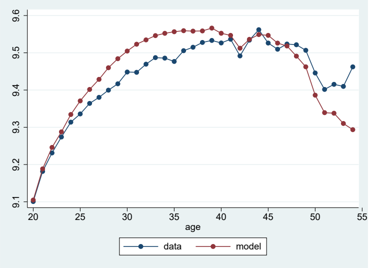
(a) By gender (b) With/without the NRPS

Figure: Fraction in the non-agricultural sector

Model fit



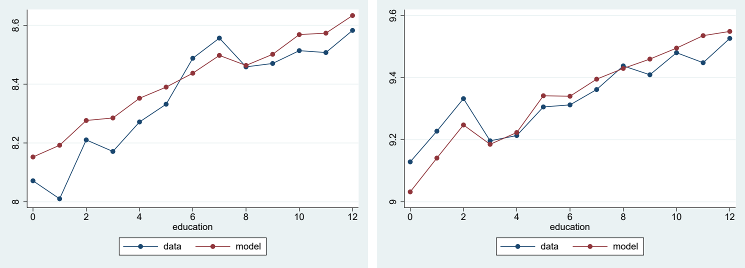
(a) Agricultural



(b) Non-agricultural

Figure: Log earnings by age

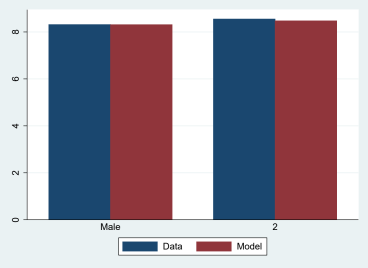
Model fit



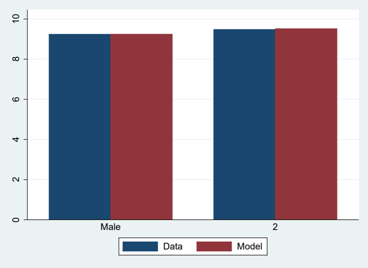
(a) Agricultural (b) Non-agricultural

Figure: Log earnings by education

Model fit



(a) Agricultural



(b) Non-agricultural

Figure: Log earnings by gender

Model fit

Table: APG

	(1) OLS Data	(2) OLS Model	(3) Individual FE Data	(4) Individual FE Model
Nonagri	0.485***	0.538***	0.516***	0.489***
	(0.015)	(0.004)	(0.017)	(0.006)
Year FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	No	No
Individual control	Yes	Ye	Yes	Yes
Individual FE	No	No	Yes	Yes
Observations	208,246	208,246	202,525	202,525

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$