Growing without Divergence: The Impact of Innovation on Low- and High-skilled Migration in China

Suqin Ge[1] Naijia Guo[2] Zibin Huang[3] Junsen Zhang[4] Li Zhang[5]

[1]VT [2]HKU [3]SHUFE [4]ZJU [5]RFBerlin

August, 2026

Motivation: Spatial Sorting and Inequality

- How workers sort into different regions has important implications for spatial inequality in developed countries (Diamond, 2016; Guerrieri, Hartley, and Hurst, 2013)
  - We typically observe a **regional divergence** where faster growing regions attracts more high-skill workers who further accelerate growth

Motivation: Spatial Sorting and Inequality

- How workers sort into different regions has important implications for spatial inequality in developed countries (Diamond, 2016; Guerrieri, Hartley, and Hurst, 2013)
  - We typically observe a **regional divergence** where faster growing regions attracts more high-skill workers who further accelerate growth
- China experienced high technological growth during the last four decades.
- China's great migration: More than 250 million Chinese people live in cities other than where they register their hukou.
- **Did China's tech growth lead to similar spatial sorting and divergence?**

Change of Skill Ratio and Initial Skill Ratio

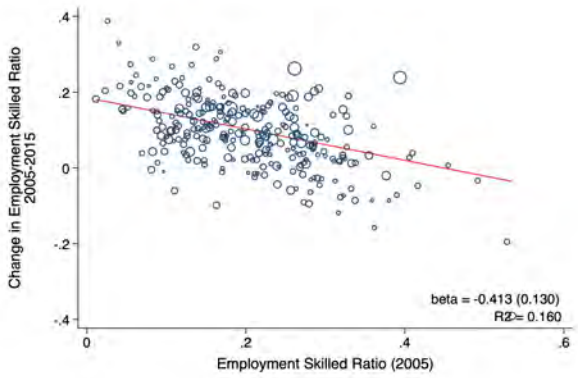


Figure 1: Initial Employment Skill Ratio and Its Changes (2005-2015)

This Paper

- We focus on the following research questions:
  - How do innovations impact migration patterns for different skilled groups in China?
  - How does spatial sorting affect welfare and inequality?
- We provide empirical evidence on the causal effect of innovation shocks on wages and migration of low- and high-skilled workers.
- We construct and structurally estimate a spatial GE model, considering:
  - Endogenous labor market, housing market, and amenity supply

Preview of Findings

- Descriptive evidence:
  - Cities with stronger patent growth experienced faster wage growth for both low- and high-skilled workers
  - However, these cities attract **more** low-skilled migrants than high-skilled.
- Structural estimation results:
  - Low-skilled workers care more about **wages**, while high-skilled workers value **amenities** more.
  - A positive shock in patents reduces the skill ratio and amenities, and discourages high-skilled migrants.
- Counterfactual of reducing patent from 2015 level to 2005 level:
  - Welfare loss for both skills, especially low-skilled people with non-agriculture hukou.
  - Welfare inequality would have been higher if technology progresses were eliminated.

Literature Review

- Spatial model and migration in China (Liang et al. 2017; Tombe and Zhu 2018; Fan 2019; Sieg, Yoon, and Zhang 2023)
  - We focus on local innovation shocks, utilize micro-level data to examine how technological development reshapes city characteristics.
- Effect of technology on employment and earnings (Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011; Acemoglu and Restrepo 2018a, 2018b, 2019, 2020, Battistii, Dustmann, and Schönberg 2023)
  - We use a GE framework to study the impact of innovation on employment, wages, and welfare in China.
- How workers sort into different local labor markets (Bayer, McMillan, and Rueben 2004; Bayer, Ferreira, and McMillan 2007; Card, Mas, and Rothstein 2008; Diamond 2016; Fajgelbaum and Gaubert 2020; Bilal and Rossi-Hansberg 2021)
  - Our results differ from those on developed countries due to technology and policy differences.

Outline of the Presentation

1. Data and Descriptive Statistics
2. Model Setting
3. Estimation
4. Counterfactual
5. Conclusion

Data and Sample

- Data:
  - Migration: micro-level census 2005, 2010, 2015
  - City-level wages by skill levels, housing prices, and amenities: statistical yearbooks
  - Patent: China National Intellectual Property Administration and Google Patent
- Sample and key definitions:
  - Unit of analysis: prefecture-level city
  - Workforce: age 25 - 50, currently working
  - Migrants: those who left hukou city for at least 6 months
  - High-skilled workers: having any college or above education
  - Sectors: urban regions + non-agricultural industries (NA) vs. rural regions (A)

Spatial Distribution of Patent Shock

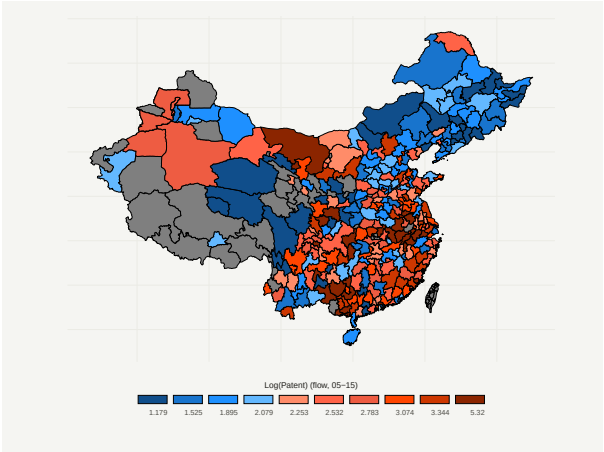


Figure 2: Spatial Distribution of Log Patent Growth (2005 - 2015)

Spatial Distribution of Innovation

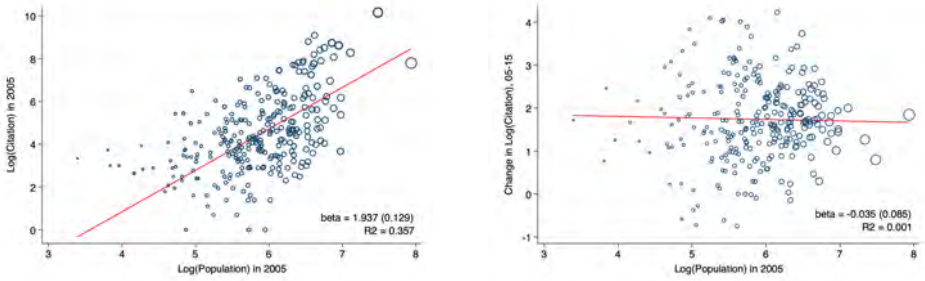


Figure 3: Citation and Its Growth and Initial Employment

Shift-share Bartik

$$\Delta P_{kt} = \sum_{ind} (Citation_{ind,-k,t} - Citation_{ind,-k,2005}) \frac{E_{ind,k,2005}}{E_{k,2005}} \quad (1)$$

- $Citation_{ind,-k,t}$: log patent citations in two-digit industry ind in year t in the country, excluding city k.
- $E_{ind,k,2005}$: employment in ind in city k in 2005.
- $E_{k,2005}$: total employment in city k in 2005.

Shift-share Bartik

$$\text{First Stage:} \quad \Delta Citation_{kt} = \alpha_0 + \alpha_1 \Delta P_{kt} + \gamma_t + \delta_k + \epsilon_{kt}$$

$$\text{Second Stage:} \quad \Delta Y_{kt} = \beta_0 + \beta_1 \Delta \widehat{Citation}_{kt} + \gamma_t + \delta_k + \epsilon_{kt}$$

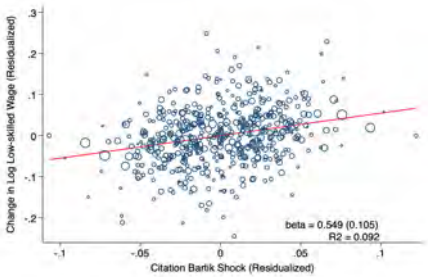
$$\text{Reduced Form:} \quad \Delta Y_{kt} = \beta_0 + \beta_1 \Delta P_{kt} + \gamma_t + \delta_k + \epsilon_{kt}$$

- ΔY_{kt}: differences from 2005 to 2010 and from 2005 to 2015
- $\Delta Citation_{kt}$: change of log number of patent citation
- ΔP_{kt}: Bartik-style citation shock

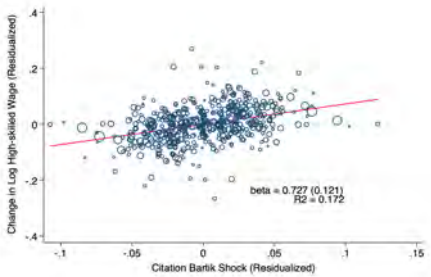
▸ Validity of Shift-Share

Empirical Analysis

Wages



(a) Low-skilled (coef = 0.549***)



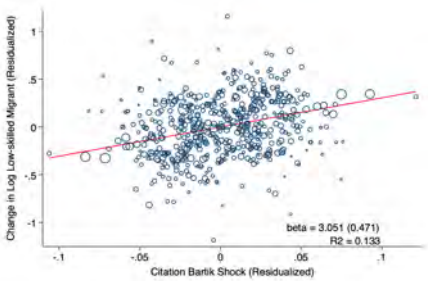
(b) High-skilled (coef = 0.727***)

Figure 4: Effect of Citation Shock on Wages for High- and Low-skilled Workers

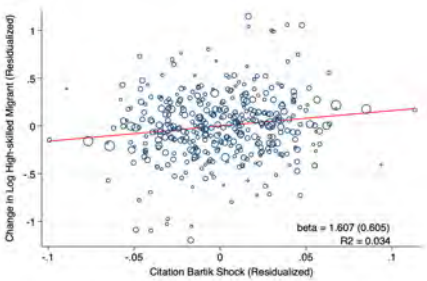
▸ Patents ▸ RoW Shift

Empirical Analysis

Migration



(a) Low-skilled (coef = 0.605***)



(b) High-skilled (coef = 0.471***)

Figure 5: Effect of Citation Shock on Number of High- and Low-skilled Migrants

▸ Patents ▸ RoW Shift

Empirical Analysis

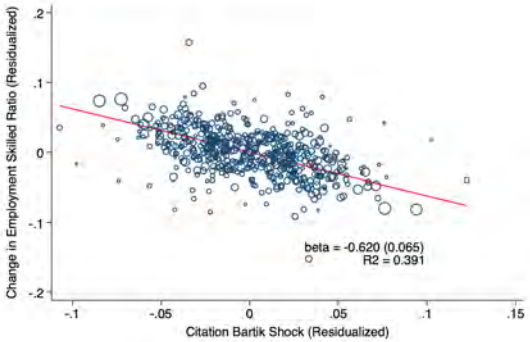


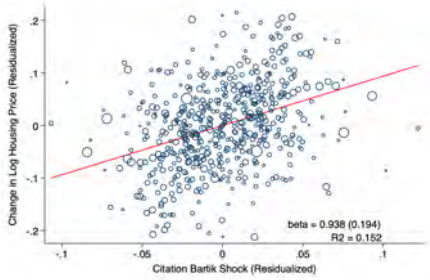
Figure 6: Citation Shock and Change in Skilled Ratio (coef = -0.620***)

▸ Patents ▸ RoW Shift ▸ Service Sector

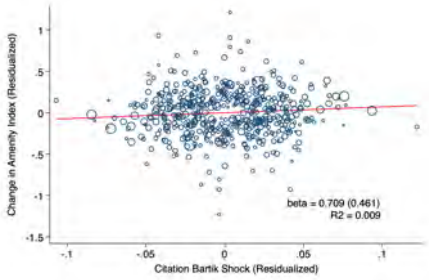
Amenity Index

	Loading	Unexplained variance
Panel A: Healthcare Index		
Hospital per 10,000 residents	0.707	0.435
Doctors per 10,000 residents	0.707	0.435
Panel B: Infrastructure Index		
Kilometers of road per 10,000 residents	0.418	0.808
Highway passengers per 10,000 residents	0.599	0.605
High-speed railway	0.683	0.486
Panel C: Environment Index		
PM 2.5	0.532	0.339
Heavily polluted days	0.571	0.237
Polluted days	0.626	0.084
Panel D: Education Index		
Teacher-student ratio in primary schools	0.082	0.982
Teacher-student ratio in middle schools	0.114	0.966
Number of colleges	0.540	0.233
Number of Project 985 universities	0.589	0.088
Number of Project 211 universities	0.586	0.097
Panel E: Amenity Index		
Healthcare Index	0.643	0.439
Infrastructure Index	0.554	0.585
Environment Index	-0.234	0.926
Education Index	0.474	0.695

Empirical Analysis



(a) Housing price (coef = 0.938***)



(b) Amenity (coef = 0.709)

Figure 7: Effect of Citation Shock on Housing Price and Amenity

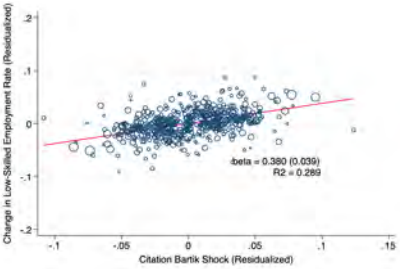
▸ Patents

Robustness Checks

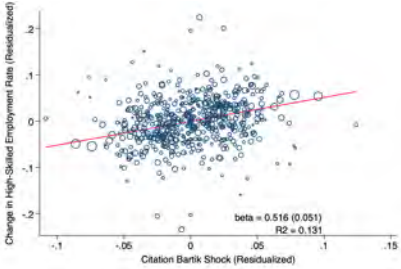
- The identification assumes that shifts are randomly assigned (Borusyak, Hull, and Jaravel 2025).
  - We show that the Bartik instrument is not correlated with city characteristics.
  - We estimate an equivalent shift-level regression.
- We use patent growth in the rest of the world as shifts.
- We rule out demand-side explanation.

Local Employment Rate

- Local employment rate: # local employment (employed workers with local hukou) / # local labor force (working age locals)



(a) Low-skilled (coef = 0.380**)



(b) High-skilled (coef = 0.516*)

Figure 8: Effect of Citation Shock on Local Employment Rate

Model Setup

- K cities in China, indexed by $k \in \{1, ..., K\}$
- Two sectors, $j \in \{a, na\}$
- Workers differ in home location k_0, hukou type j_0, and skill $e \in \{L, H\}$.
- Each worker i first chooses the sector j, then chooses which city k to live
- Non-ag hukou workers can only choose the non-ag sector
- Three markets: labor market, housing market, and amenity market
- Labor supply, housing supply, and amenity supply are endogenous in the model
- Patents affect Non-ag productivity and amenity

Labor Demand — Non-agricultural Sector

- NA Firms produce a homogeneous tradable good
- Using technology A_{kt}, high-skilled labor $H_{na,kt}$, low-skilled labor $L_{na,kt}$, capital $K_{na,kt}$, and machine $C_{na,kt}$

$$
\begin{aligned}
Y_{na,kt} &= z_{na,kt} N_{na,kt}^{\alpha} (\theta_{kt}^{K} K_{kt})^{1-\alpha} \\
N_{na,kt} &= (\theta_{kt}^{L} (L_{na,kt} + \omega C_{kt})^{\rho} + \theta_{kt}^{H} H_{na,kt}^{\rho})^{\frac{1}{\rho}} \\
C_{kt} &= f_C(A_{kt}) \\
\theta_{kt}^{K} &= f_K(A_{kt}) \\
\theta_{kt}^{L} &= f_L(A_{kt}, H_{na,kt}, L_{na,kt}), \theta_{kt}^{H} = f_L(A_{kt}, H_{na,kt}, L_{na,kt})
\end{aligned}
$$

- L and C are substitutes. θ is factor-augmenting productivity

Labor Demand

- Using log-linear transformation, we get:

$$w^H_{na,kt} = \gamma_{HA} A_{kt} + \gamma_{na,HH} \ln(H_{na,kt}) + \gamma_{na,HL} \ln(L_{na,kt}) + d^H_{na,kt}$$
$$w^L_{na,kt} = \gamma_{LA} A_{kt} + \gamma_{na,LH} \ln(H_{na,kt}) + \gamma_{na,LL} \ln(L_{na,kt}) + d^L_{na,kt}$$

- Labor demand function in agricultural sector is:

$$w^H_{ag,kt} = w^L_{ag,kt} = \gamma_{ag} \ln(H_{ag,kt} + L_{ag,kt}) + d_{ag,kt}$$

Labor Supply: Location + Sector Choices

- The skill-specific utility of working in the non-ag sector in city k and year t is

$$V_{ikt} = \beta_1^e w_{na,kt}^e + \beta_2^e r_{kt} + \beta_3^e a_{kt} + MigrationCost_{ikt} + \nu_{kt}^e + \epsilon_{ikt}$$

- Migration cost can be decomposed into:

$$MigrationCost_{ikt} = \sum_r \beta_{4rt}^e WithinHometown_{ikt} 1_{k \in r} + \sum_\tau \beta_{\tau 5 lt}^e WithinProvince_{ikt} 1_{k \in \tau} + \beta_{6t}^e hukou_{ikt} + \beta_{7t}^e hukou_{ikt}^2$$

- The value of working in the agricultural sector in prefecture k is

$$W_{it}^a = \alpha_{0k_0} + \alpha_1 w_{a,k_0 t}^e + \xi_{it}^a$$

Housing and Amenity Supply

- The housing supply equation is

$$r_{kt} = \ln(R_{kt}) = \ln(\iota_t) + \ln(CC_{kt}) + \left[\gamma_1^{hd} + \gamma_2^{hd} geo_k\right] \ln(HD_{kt})$$
$$HD_{kt} = L_{na,kt} W_{na,kt}^{L} + H_{na,kt} W_{na,kt}^{H}$$

- Endogenous amenity depends on technology and skill ratio:

$$a_{kt} = \gamma_1^a A_{kt} + \gamma_2^a \ln\left(\frac{H_{na,kt}}{L_{na,kt}}\right) + \epsilon_{kt}^a$$

  - Amenities respond to the education of neighbours (Bayer, Ferreira, and McMillan, 2007) and college employment ratios (Diamond, 2016).
  - Patents have a direct impact on amenities.

Estimation Results

- The labor supply model is solved by BLP approach
- For the rest, we take the first difference of these equations, then use different Shift-Share IVs to estimate the first-differences
  - IV for labor supply: migration Bartik (Card, 2009)
  - IV for housing demand: wage Bartik (Diamond, 2016)
  - IV for skilled ratio: employment Bartik
- The model fit is good ▸ Model Fit

Table 1: Estimation of the Labor Demand in the Non-agricultural Sector

VARIABLES	(1) Δ Log High-skilled Wage	(2) Δ Log Low-skilled Wage	(3) Δ Log High-skilled Employ-ment	(4) Δ Log Low-skilled Employ-ment	(5) Δ Log Citation	(6) Δ Log High-skilled Wage	(7) Δ Log Low-skilled Wage
Δ Log Citation	0.033***	0.033**				1.099***	1.036***
	(0.012)	(0.013)				(0.316)	(0.302)
Δ Log High-skilled Employment	0.040*	0.041				-0.078	0.08
	(0.024)	(0.027)				(0.501)	(0.479)
Δ Log Low-skilled Employment	-0.061*	-0.067*				-0.726	-0.848
	(0.031)	(0.035)				(0.719)	(0.688)
Citation shock			-0.190	0.110	0.566**		
			(0.137)	(0.108)	(0.240)		
Migrant bartik for high-skill workers			0.627***	0.409***	0.344		
			(0.143)	(0.113)	(0.250)		
Migrant bartik for low-skill workers			-0.228	0.614***	-0.182		
			(0.255)	(0.201)	(0.446)		
Observations	468	468	468	468	468	468	468
R-squared	0.025	0.021					
Model	OLS	OLS	First stage	First stage	First stage	IV GMM	IV GMM
Cragg-Donald Wald F						3.510	3.510
Sanderson-Windmeijer F			15.32	16.28	11.22		

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Estimation of the Housing Market

VARIABLES	(1) Δ Log(Rent)	(2) Δ Log Housing Demand	(3) Δ Log Housing Demand * Geo	(4) Δ Log(Rent)
Δ Log housing demand	0.0118			0.599***
	(0.0663)			(0.120)
Δ Log housing demand * Log altitude	0.0290**			0.0450***
	(0.0116)			(0.0167)
Wage Shift-Share IV for High-skilled Workers		-5.737***	-18.66***	
		(1.235)	(6.272)	
Wage Shift-Share IV for Low-skilled Workers		6.693***	17.16***	
		(1.226)	(6.228)	
Wage Shift-Share IV for High-skilled Workers * Log altitude		0.956***	3.284***	
		(0.211)	(1.069)	
Wage Shift-Share IV for Low-skilled Workers * Log altitude		-0.948***	-2.046*	
		(0.210)	(1.065)	
Observations	468	468	468	468
R-squared	0.045			
Model	OLS	First stage	First stage	IV GMM
Cragg-Donald Wald F				41.79
Sanderson-Windmeijer F		607.3	93.77	

Standard errors in parentheses

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

Table 3: Estimation of the Amenity Market

VARIABLES	(1) Δ Amenity Index	(2) Δ Log High-skilled Employment Ratio	(3) Δ Log Citation	(4) Δ Amenity Index
Δ Log Citation	0.152***			1.084***
	(0.034)			(0.309)
Δ High-skilled Employment Ratio	0.492			5.769**
	(0.387)			(2.345)
Citation shock		-0.081***	0.426*	
		(0.022)	(0.251)	
Employment Shift-Share IV for high-skilled workers		0.077	-1.108	
		(0.059)	(0.678)	
Employment Shift-Share IV for low-skilled workers		0.079	1.573**	
		(0.062)	(0.722)	
Observations	468	468	468	468
R-squared	0.044			
Model	OLS	First stage	First stage	IV GMM
Cragg-Donald Wald F				15.4
Sanderson-Windmeijer F		14.28	8.193	

Standard errors in parentheses

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

Table 4: Estimation of Workers' Location Choice (Second Stage)

VARIABLES	(1) $\Delta\delta_{high}$	(2) $\Delta\delta_{low}$	(3) $\Delta\delta_{high}$	(4) $\Delta\delta_{low}$
Δ Log(High-skilled Wage)	0.328**		1.161**	
	(0.157)		[0.583]	
Δ Log(Low-skilled Wage)		0.322		1.673**
		(0.227)		[0.700]
Δ Log(Housing Rent)	0.148	0.165	-0.929*	-0.954*
	(0.101)	(0.128)	[0.511]	[0.504]
Δ Amenity Index	0.0656	0.0733	0.371**	0.196
	(0.0524)	(0.0615)	[0.164]	[0.266]
Constant	0.447***	0.557***	0.471***	0.358
	(0.125)	(0.167)	[0.179]	[0.260]
Observations	476	476	451	451
R-squared	0.035	0.033	0.055	0.058
Model	OLS	OLS	IV	IV

Mechanism

- Low-skilled workers:
  - Patent $\uparrow \Rightarrow$ wage $\uparrow \Rightarrow$ migration $\uparrow$ (Important)
  - Patent $\uparrow \Rightarrow$ amenity $\uparrow \Rightarrow$ migration $\uparrow$ (Less Important)

Mechanism

- Low-skilled workers:
  - Patent $\uparrow \Rightarrow$ wage $\uparrow \Rightarrow$ migration $\uparrow$ (Important)
  - Patent $\uparrow \Rightarrow$ amenity $\uparrow \Rightarrow$ migration $\uparrow$ (Less Important)
- High-skilled workers:
  - Patent $\uparrow \Rightarrow$ wage $\uparrow \Rightarrow$ migration $\uparrow$ (Less Important)
  - Patent $\uparrow \Rightarrow$ amenity $\uparrow \Rightarrow$ migration $\uparrow$ (Important)
  - Patent $\uparrow \Rightarrow$ skilled-ratio $\downarrow \Rightarrow$ amenity $\downarrow \Rightarrow$ migration $\downarrow$ (Important)

Counterfactual Analysis

1. Exploring the impacts of technological growth:
   Fix patent to the 2005 level
2. Disentangling difference channels:
   Shut down the feedback of wages, rents, and amenities in turn
3. Policy counterfactual:
   Equalize the preference for amenities for different skill groups

Counterfactual I: Patents Reduce to 2005 Level

	Original Eq	Counterfactual	Change
Panel A. Migration across Cities			
Total Migration	26644290	19039122	-28.54%
High-skilled Migration	5746083	5499044	-4.30%
Low-skilled Migration	20898208	13540079	-35.21%
Panel B. Urban Skill Ratio and Workforce Ratio			
National Urban Skill Ratio	0.360	0.421	16.94%
National Urban Ratio	0.412	0.335	-18.69%
Panel C. Average Wage by Skill and Sector (RMB)			
Wage of Low-skilled in Agr	12650.54	12351.34	-2.37%
Wage of Low-skilled in Non-agr	47076.32	14186.43	-69.87%
Wage of High-skilled in Agr	12650.54	12351.34	-2.37%
Wage of High-skilled in Non-agr	57090.14	15961.89	-72.04%
Panel D. Inequality			
National Wage Gini Coefficient	0.430	0.253	-41.16%
National Welfare Gini Coefficient	0.0965	0.103	6.74%

Counterfactual II: Channel Decomposition

- In addition to fixing patent to the 2005 level, we further fix each of the following to the 2015 level when estimating the counterfactual equilibrium:
  1. Wage
  2. Rent
  3. Amenities that are determined by innovation
  4. Amenities that are determined by the skill ratio

Counterfactural II: Channel Decomposition

- This table shows the coefficient of regressing migration on patent shock

	High-skilled Migration	Low-skilled Migration
Original Counterfactual	0.438***	0.583***
	(0.012)	(0.016)
Fixed Wage	0.309***	0.136***
	(0.005)	(0.003)
Fixed Rent	0.970***	1.059***
	(0.018)	(0.025)
Fixed Direct Amenity	0.135***	0.465***
	(0.010)	(0.013)
Fixed Skill Ratio	0.481***	0.600***
	(0.012)	(0.016)

Counterfactual III: Equalizing Amenity Preferences

- In this exercise, we take technological growth as given, but set low-skilled workers' preference for amenity (β_3^L) to be the same as high-skilled workers (β_3^H)
- It mimics a *hukou* reform that equalizes access to public goods across skill groups

Counterfactual III: Equalizing Amenity Preferences

	Original Eq	Counterfactual	Change
Panel A. Migration across Cities			
Total Migration	26644290	31291350	17.44%
High-skilled Migration	5746082.5	5297753	-7.80%
Low-skilled Migration	20898208	25993596	24.38%
Panel B. Urban Skill Ratio and Workforce Ratio			
National Skill Ratio	0.360	0.331	-8.02%
National Urban Ratio	0.412	0.444	7.74%
Panel C. Wage and Welfare Inequality			
Wage Gini Coefficient	0.430	0.393	-8.6%
Welfare Gini Coefficient	0.0965	0.0910	-5.7%

Conclusion

- Innovation in China during 2005 to 2015 increased wages for both low- and high-skilled workers
- Low-skilled workers care more about wages while high-skilled workers care more about amenities
- Innovation attracted more low-skilled workers, reduced the skill ratio, which reduced amenities and discourages high-skilled migrants
- Technology growth in China during the last twenty years DID NOT lead to a divergence and benefited workers with different skills

Summary Statistics

Back

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Migrant/Employment	609	0.16	0.16	0.00	0.90
High-Skilled Migrant/High-Skilled Employment	609	0.08	0.09	0.00	0.59
Low-Skilled Migrant/Low-Skilled Employment	609	0.19	0.19	0.00	0.96
Citations of Patents	571	2903.20	11265.63	2.00	157306.00
High-skilled wage	595	48127.27	14217.58	15928.67	122615.09
Low-skilled wage	595	39474.81	11309.51	6007.17	91138.81
City-level average house price	570	4822.948	3086.891	1589.353	33942.34
Doctors per 10,000 residents	576	20.37	8.18	6.92	75.19
Hospitals per 10,000 residents	576	0.60	0.65	0.09	6.89
Kilometers of road per 10,000 residents	575	33.38	18.72	1.44	152.09
Highway passengers per 10,000 residents	574	24.30	121.30	1.15	2855.72
High-speed Railway Connected	577	0.41	0.49	0.00	1.00
PM 2.5	572	44.64	20.05	4.15	101.19
Heavily polluted days	576	6.85	10.91	0.00	55.89
Polluted days	576	70.37	56.32	0.00	237.05
Teacher-student ratio in primary schools	577	0.06	0.01	0.00	0.13
Teacher-student ratio in middle schools	576	0.08	0.02	0.00	0.20
Number of colleges	565	8.45	14.55	1.00	90.00
Number of Project 985 universities	578	0.13	0.64	0.00	8.00
Number of Project 211 universities	578	0.38	1.75	0.00	23.00
Log altitude	574	5.46	1.50	0.26	8.48

Descriptive Analysis

Back

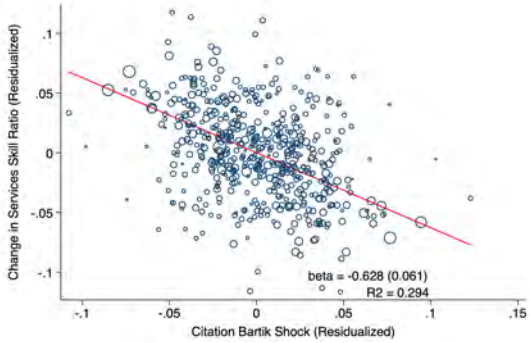


Figure 9: Citation Shock and Change in Skilled Ratio in Service Sector

Descriptive Analysis

Back

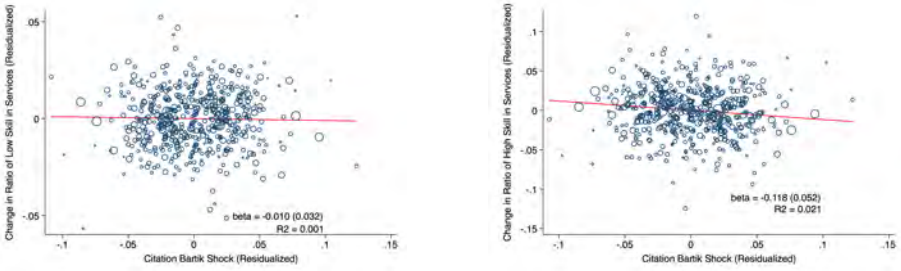


Figure 10: Citation Shock and High- and Low-Skilled Workers Choosing Service Sector

Patent Shift-Share Shock

$$\Delta P_{kt} = \sum_{ind} (Patent_{ind,-k,t} - Patent_{ind,-k,2005}) \frac{E_{ind,k,2005}}{E_{k,2005}}$$

- $Patent_{ind,-k,t}$ represents the log number of patent in industry *ind* in year t in the country, excluding city k.
- $E_{ind,k,2005}$ and $E_{k,2005}$ measure the number of workers in industry *ind* in city k in year 2005.
- Measure exogenous technology shocks
- Industry: two-digit, total 95 industries (54 with patent)

Back to Descriptive | Back to Estimation

Descriptive Analysis Using Patents

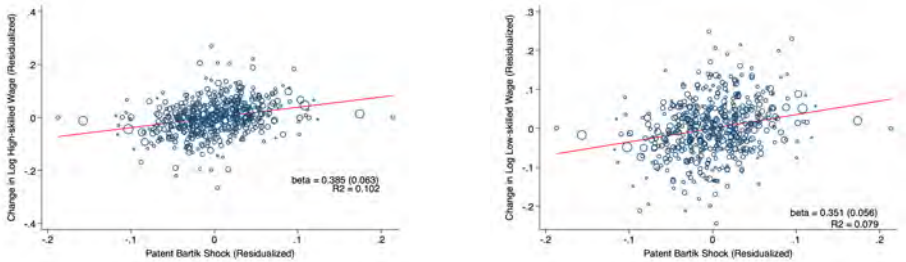


Figure 11: Effect of Patent Shock on Wages for High- and Low-skilled Workers

Back

Descriptive Analysis Using Patents

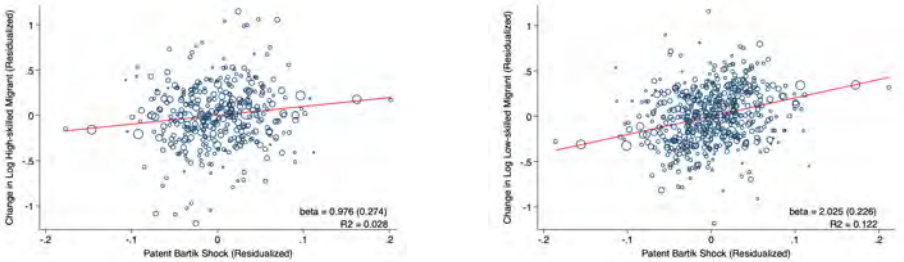


Figure 12: Effect of Patent Shock on Number of High- and Low-skilled Migrants

Back

Descriptive Analysis Using Patents

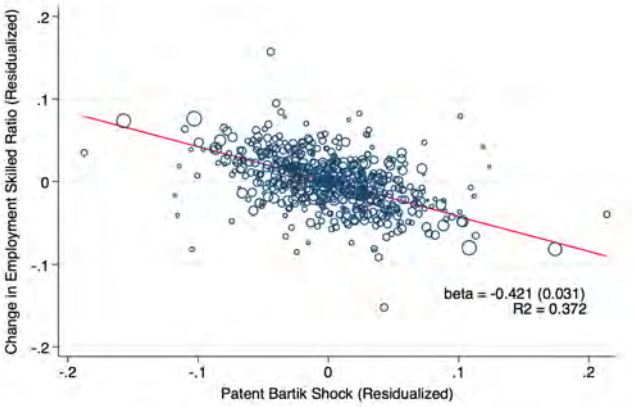


Figure 13: Patent Shock and Change in Skilled Ratio

Descriptive Analysis Using Patents

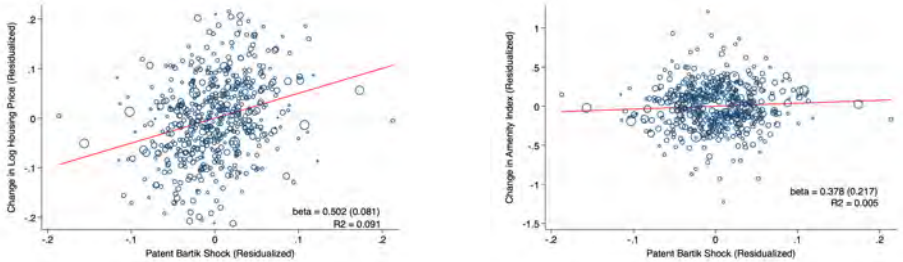


Figure 14: Effect of Patent Shock on Housing Price and Amenity

Back

Patent growth and labor supply

VARIABLES	(1) Δ Log Employment	(2) Δ Log High-Skilled Employment	(3) Δ Log Low-Skilled Employment	(4) Δ Log High-skilled Migrants	(5) Δ Log Low-skilled Migrants	(6) Δ Employment Skilled Ratio
Panel A: OLS						
Δ Log(Patent)	0.0903**	0.0157	0.127**	0.114	0.192***	-0.0267***
	(0.0397)	(0.0362)	(0.0509)	(0.0817)	(0.0724)	(0.00993)
Panel B: Reduced Form						
Patent shock	1.256***	-0.0877	2.007***	0.976**	2.025***	-0.421***
	(0.235)	(0.199)	(0.242)	(0.389)	(0.320)	(0.0435)
Panel C: IV						
Δ Log(Patent)	1.103***	-0.111	1.795***	0.675**	1.874***	-0.373***
	(0.302)	(0.187)	(0.442)	(0.312)	(0.540)	(0.0888)
Year FE	X	X	X	X	X	X
City FE	X	X	X	X	X	X

Back

Patent growth, wage, housing price, and amenity

VARIABLES	(1) Δ Log High-skilled Wage	(2) Δ Log Low-skilled Wage	(3) Δ Log(Housing Price)	(4) Δ Amenity Index
Panel A: OLS				
Δ Log(Patent)	-0.00123	-0.00769	-0.0249	0.0954
	(0.0162)	(0.0174)	(0.0198)	(0.0799)
Panel B: Reduced Form				
Patent shock	0.385***	0.351***	0.502***	0.378
	(0.0887)	(0.0795)	(0.115)	(0.307)
Panel C: IV				
Δ Log(Patent)	0.402***	0.348***	0.469**	0.428
	(0.146)	(0.134)	(0.199)	(0.306)
Year FE	X	X	X	X
City FE	X	X	X	X

Back

Tests of Shift-Share Shocks

Table 5: Balance Test of Shift-Share Shock of Citation

VARIABLES	2010		2015	
	Coefficient	S.E.	Coefficient	S.E.
GDP Growth Rate	0.00150	(0.00303)	0.000899	(0.00244)
Log Population Density	0.00166	(0.00871)	-0.00254	(0.00493)
Log Fiscal Expenditure	0.0168	(0.0226)	-0.00361	(0.0283)
Log Retail Sale	-0.00131	(0.0128)	-0.0140	(0.00871)
Log Above-Scale Firm Profit	-0.0956	(0.0868)	0.0287	(0.108)
Log Number of College Students	-0.00429	(0.0204)	-0.00376	(0.0200)

Back

Descriptive Analysis Using RoW Shifts

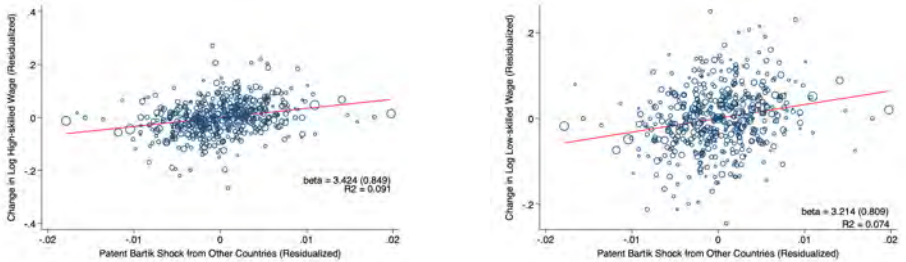


Figure 15: Effect of RoW Patent Shock on Wages for High- and Low-skilled Workers

Back

Descriptive Analysis Using RoW Shifts

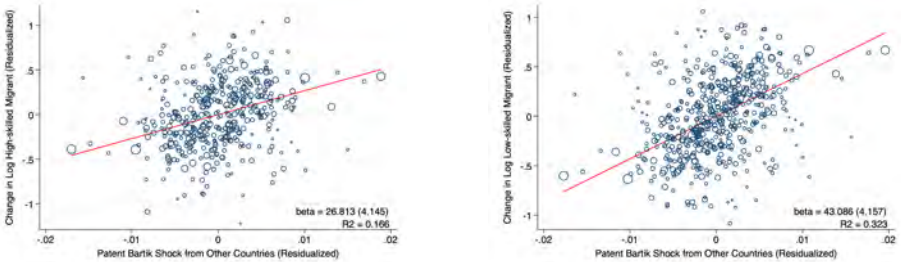


Figure 16: Effect of RoW Patent Shock on High- and Low-skilled Migrants

Back

Descriptive Analysis Using RoW Shifts

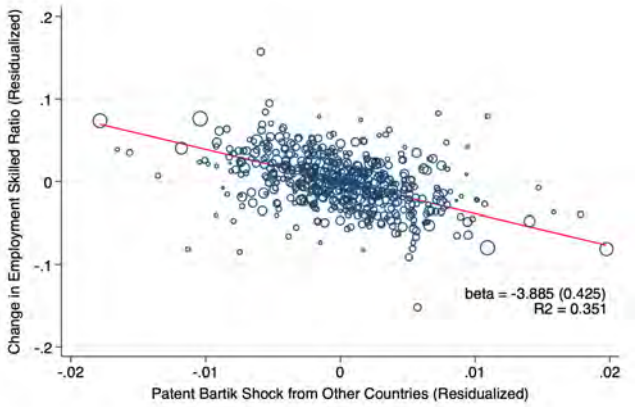


Figure 17: Effect of RoW Patent Shock on Skill Ratio

Estimation of Labor Demand in the Non-agricultural Sector

Back

- We use first difference regression to estimate labor demand equations

$$\Delta w^{H}_{na,kt} = \gamma_{HA}\Delta A_{kt} + \gamma_{na,HH}\Delta \ln H_{na,kt} + \gamma_{na,HL}\Delta \ln L_{na,kt} + \Delta \epsilon^{H}_{na,kt}$$
$$\Delta w^{L}_{na,kt} = \gamma_{LA}\Delta A_{kt} + \gamma_{na,LH}\Delta \ln H_{na,kt} + \gamma_{na,LL}\Delta \ln L_{na,kt} + \Delta \epsilon^{L}_{na,kt}$$

- Need variation in labor supply uncorrelated with unobserved changes in local productivity
- Instrument for low- and high-skilled workforce: migrant shift-share shock
- Instrument for local patent growth: patent shift-share shock

Migrant Shift-Share Shock

Back

- National changes in the number of migrants at industry level $\times$ share of migrants in each industry in a location (Card, 2009):

$$\Delta B^{e}_{kt} = \sum_{\text{ind}} \left(Mig^{e}_{\text{ind},na,-k,t} - Mig^{e}_{\text{ind},na,-k,2005} \right) \frac{Mig^{e}_{ind,na,k,2005}}{Mig^{e}_{na,k,2005}}$$

- $Mig^{e}_{ind,na,-k,t}$: total number of skill e migrants in industry ind in year t, excluding migrants in city k
- $Mig^{e}_{ind,na,k,2005}$: number of skill e migrants in industry ind in city k in year 2005.

Tests of Shift-Share Shocks

Table 6: Balance Test of Shift-Share Shock of Migrants

VARIABLES	2010 Coefficient	S.E.	2015 Coefficient	S.E.
Panel A: High-Skilled				
GDP Growth Rate	-0.00278	(0.00673)	0.00250	(0.00520)
Log Population Density	-0.00762	(0.00779)	-0.0202*	(0.0108)
Log Fiscal Expenditure	-0.0415	(0.0573)	-0.0165	(0.0942)
Log Retail Sale	0.0195	(0.0496)	-0.0145	(0.0665)
Log Above-Scale Firm Profit	-0.0109	(0.118)	0.0746	(0.106)
Log Number of College Students	0.725	(0.476)	0.258	(0.559)
Panel B: Low-Skilled				
GDP Growth Rate	-0.00629**	(0.00230)	-0.00404	(0.00298)
Log Population Density	0.00242	(0.00196)	0.00609	(0.00486)
Log Fiscal Expenditure	-0.0382	(0.0249)	-0.0357	(0.0395)
Log Retail Sale	-0.00912	(0.0123)	-0.0432*	(0.0232)
Log Above-Scale Firm Profit	-0.0790	(0.0683)	0.0483	(0.138)
Log Number of College Students	0.403	(0.245)	0.143	(0.261)

Back

Table 7: Estimation of the Labor Demand in the Non-agricultural Sector

VARIABLES	(1) Δ Log High-skilled Wage	(2) Δ Log Low-skilled Wage	(3) Δ Log High-skilled Employ-ment	(4) Δ Log Low-skilled Employ-ment	(5) Δ Log Citation	(6) Δ Log High-skilled Wage	(7) Δ Log Low-skilled Wage
Δ Log Citation	0.033***	0.033**				1.099***	1.036***
	(0.012)	(0.013)				(0.316)	(0.302)
Δ Log High-skilled Employment	0.040*	0.041				-0.078	0.08
	(0.024)	(0.027)				(0.501)	(0.479)
Δ Log Low-skilled Employment	-0.061*	-0.067*				-0.726	-0.848
	(0.031)	(0.035)				(0.719)	(0.688)
Citation shock			-0.190	0.110	0.566**		
			(0.137)	(0.108)	(0.240)		
Migrant bartik for high-skill workers			0.627***	0.409***	0.344		
			(0.143)	(0.113)	(0.250)		
Migrant bartik for low-skill workers			-0.228	0.614***	-0.182		
			(0.255)	(0.201)	(0.446)		
Observations	468	468	468	468	468	468	468
R-squared	0.025	0.021					
Model	OLS	OLS	First stage	First stage	First stage	IV GMM	IV GMM
Cragg-Donald Wald F						3.510	3.510
Sanderson-Windmeijer F			15.32	16.28	11.22		

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Back Patents

Estimation of the Labor Demand in the NA Sector (Patent)

Table 8: Estimation of the Labor Demand in the Non-agricultural Sector

VARIABLES	(1) Δ Log High-skilled Wage	(2) Δ Log Low-skilled Wage	(3) Δ Log High-skilled Employment	(4) Δ Log Low-skilled Employment	(5) Δ Log Patent	(6) Δ Log High-skilled Wage	(7) Δ Log Low-skilled Wage
Δ Log Patent (Flow, IPA)	0.114***	0.114***				0.514***	0.481***
	(0.0116)	(0.0132)				(0.0877)	(0.0795)
Δ Log High-skilled Employment	0.00306	0.00559				-0.777*	-0.556
	(0.0224)	(0.0254)				(0.435)	(0.394)
Δ Log Low-skilled Employment	-0.0584**	-0.0635**				0.107	-0.0860
	(0.0282)	(0.0319)				(0.426)	(0.387)
Patent shock			-0.0207	0.100	0.978***		
			(0.125)	(0.0989)	(0.175)		
Migrant bartik for high-skill workers			0.564***	0.404***	0.659***		
			(0.142)	(0.112)	(0.199)		
Migrant bartik for low-skill workers			-0.162	0.721**	-0.253		
			(0.357)	(0.282)	(0.500)		
Constant	0.543***	0.545***	-0.226	-0.832***	-0.371	0.185*	0.174*
	(0.0233)	(0.0263)	(0.348)	(0.275)	(0.488)	(0.0980)	(0.0889)
Observations	484	484	484	484	484	484	484
R-squared	0.174	0.142					
Model	OLS	OLS	First stage	First stage	First stage	IV GMM	IV GMM
Cragg-Donald Wald F						2.266	2.266
Sanderson-Windmeijer F			6.819	9.238	9.247		

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Back

Estimation of Labor Demand in the Agricultural Sector

Back

$$\Delta w_{a,kt} = \gamma_a \Delta \ln(H_{a,kt} + L_{a,kt}) + \Delta \epsilon_{a,kt}$$

- Count all workers in the rural area as agricultural workers.
- Instrument for agricultural employment: population with agricultural hukou

Table 9: Estimation of the Labor Demand in the Agricultural Sector

VARIABLES	(1) Δ Log Agr Income	(2) Δ Log Agr Employment	(3) Δ Log Agr Income
Δ log agricultural employment	-0.044*		-0.172***
	(0.024)		(0.034)
Δ log agricultural population		1.167***	
		(0.053)	
Observations	468	468	468
R-squared	0.007		
Model	OLS	First stage	IV GMM
Cragg-Donald Wald F			488.9
Sanderson-Windmeijer F		488.9	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Back

Estimation of the Housing Market

Back

$$\Delta r_{kt} = \left[\gamma_1^{hd} + \gamma_2^{hd} \times \ln(Altitude_k)\right] \Delta \ln(HD_{kt}) + \Delta \epsilon_{kt}^{r}$$

- Need variation in housing demand unrelated to changes in unobserved factors driving housing prices $\Delta \ln(CC_{kt})$
- Instrument for housing demand: wage shift-share shock

Wage Shift-Share Shock

Back

- We interact cross-sectional differences in industrial employment composition with national changes in industry wage levels to construct the wage shift-share shock following (Diamond, 2016):

$$\Delta W_{kt}^{H} = \sum_{ind} (w_{ind,na,-k,t}^{H} - w_{ind,na,-k,2005}^{H}) \frac{H_{ind,na,k,2005}}{H_{na,k,2005}}$$

$$\Delta W_{kt}^{L} = \sum_{ind} (w_{ind,na,-k,t}^{L} - w_{ind,na,-k,2005}^{L}) \frac{L_{ind,na,k,2005}}{L_{na,k,2005}}$$

- $w_{ind,na,-k,t}^{e}$: log wage of high/low skill workers in industry *ind* in year t, excluding city k
- $H_{ind,na,k,2005}$ and $L_{ind,na,k,2005}$: high/low skill workers in industry *ind* in city k in the urban area in 2005

Tests of Shift-Share Shocks

Table 10: Balance Test of Shift-Share Shock of Wages

VARIABLES	2010 Coefficient	S.E.	2015 Coefficient	S.E.
Panel A: High-Skilled				
GDP Growth Rate	-0.00591	(0.00836)	-0.00479	(0.00536)
Log Population Density	0.00448	(0.0120)	-0.0149	(0.0133)
Log Fiscal Expenditure	0.0220	(0.0703)	-0.0140	(0.117)
Log Retail Sale	-0.00487	(0.0299)	-0.0370	(0.0660)
Log Above-Scale Firm Profit	-0.114	(0.273)	0.144	(0.289)
Log Number of College Students	-1.807**	(0.683)	-2.081***	(0.688)
Panel B: Low-Skilled				
GDP Growth Rate	0.0124	(0.0110)	0.00377	(0.00721)
Log Population Density	-0.00680	(0.00533)	-0.0121*	(0.00609)
Log Fiscal Expenditure	-0.00151	(0.0765)	-0.0493	(0.113)
Log Retail Sale	-0.0129	(0.0363)	-0.0468	(0.0525)
Log Above-Scale Firm Profit	-0.212	(0.220)	0.102	(0.377)
Log Number of College Students	-0.672*	(0.327)	-0.653	(0.444)

Back

Table 11: Estimation of the Housing Market

VARIABLES	(1) Δ Log(Rent)	(2) Δ Log Housing Demand	(3) Δ Log Housing Demand * Geo	(4) Δ Log(Rent)
Δ Log housing demand	0.0118			0.599***
	(0.0663)			(0.120)
Δ Log housing demand * Log altitude	0.0290**			0.0450***
	(0.0116)			(0.0167)
Wage Shift-Share IV for High-skilled Workers		-5.737***	-18.66***	
		(1.235)	(6.272)	
Wage Shift-Share IV for Low-skilled Workers		6.693***	17.16***	
		(1.226)	(6.228)	
Wage Shift-Share IV for High-skilled Workers * Log altitude		0.956***	3.284***	
		(0.211)	(1.069)	
Wage Shift-Share IV for Low-skilled Workers * Log altitude		-0.948***	-2.046*	
		(0.210)	(1.065)	
Observations	468	468	468	468
R-squared	0.045			
Model	OLS	First stage	First stage	IV GMM
Cragg-Donald Wald F				41.79
Sanderson-Windmeijer F		607.3	93.77	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Back

Estimation of the Amenity Market

Back

$$\Delta a_{kt} = \gamma_1^a \Delta A_{kt} + \gamma_2^a \Delta \ln \left(\frac{H_{na,kt}}{L_{na,kt}} \right) + \Delta \epsilon_{kt}^a$$

- Instrument for changes in the skilled ratio: employment shift-share shock for low- and high-skilled workers
- Instrument for patent growth: patent shift-share shock

Employment Shift-Share Shock

Back

$$\Delta E^H_{kt} = \sum_{ind} (H_{ind,na,-k,t} - H_{ind,na,-k,2005}) \frac{H_{ind,na,k,2005}}{H_{na,k,2005}}$$

$$\Delta E^L_{kt} = \sum_{ind} (L_{ind,na,-k,t} - L_{ind,na,-k,2005}) \frac{L_{ind,na,k,2005}}{L_{na,k,2005}}$$

- $H_{ind,na,-k,t}$ represents the log number of high-skilled worker in industry *ind* in year t in the country, excluding city k.
- $L_{ind,na,-k,t}$ represents the log number of low-skilled worker in industry *ind* in year t in the country, excluding city k.
- $H_{na,-k,t}$, $L_{na,-k,t}$ follow the same definitions as in the construction of wage bartik.

Table 12: Estimation of the Amenity Market

VARIABLES	(1) Δ Amenity Index	(2) Δ Log High-skilled Employment Ratio	(3) Δ Log Citation	(4) Δ Amenity Index
Δ Log Citation	0.152***			1.084***
	(0.034)			(0.309)
Δ High-skilled Employment Ratio	0.492			5.769**
	(0.387)			(2.345)
Citation shock		-0.081***	0.426*	
		(0.022)	(0.251)	
Employment Shift-Share IV for high-skilled workers		0.077	-1.108	
		(0.059)	(0.678)	
Employment Shift-Share IV for low-skilled workers		0.079	1.573**	
		(0.062)	(0.722)	
Observations	468	468	468	468
R-squared	0.044			
Model	OLS	First stage	First stage	IV GMM
Cragg-Donald Wald F				15.4
Sanderson-Windmeijer F		14.28	8.193	

Standard errors in parentheses

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

Labor Supply: Location Choices

BLP model:

$$V_{ikt} = \delta^e_{kt} + MigrationCost_{ikt} + \epsilon_{ikt}$$
$$\delta^e_{kt} = \beta^e_1 w^e_{na,kt} + \beta^e_2 r_{kt} + \beta^e_3 a_{kt} + \nu^e_{kt}$$

- First step: maximum likelihood, estimate δ^e_{kt} and migration cost
- Second step:

$$\Delta\delta^e_{kt} = \beta^e_1 \Delta w^e_{na,kt} + \beta^e_2 \Delta r_{kt} + \beta^e_3 \Delta a_{kt} + \Delta\nu^e_{kt}$$

  - Instruments: Lasso-selected from a large set of potential IVs (Chernozhukov et al. 2018).

Back

Table 13: Estimation of Workers' Location Choice (First Stage)

Worker Type	Low-Skilled	Low-Skilled	Low-Skilled	High-Skilled	High-Skilled	High-Skilled
Year	2005	2010	2015	2005	2010	2015
Within Hometown (East)	4.725***	3.817***	4.23***	3.832***	3.383***	3.766***
	(0.037)	(0.016)	(0.024)	(0.042)	(0.022)	(0.027)
Within Hometown (Middle)	4.892***	4.526***	4.588***	4.644***	3.94***	4.144***
	(0.043)	(0.021)	(0.033)	(0.088)	(0.037)	(0.049)
Within Hometown (West)	4.627***	3.611***	4.015***	3.987***	3.266***	3.736***
	(0.048)	(0.022)	(0.04)	(0.077)	(0.038)	(0.056)
Within Hometown (North East)	4.628***	4.223***	4.941***	4.98***	3.873***	4.39***
	(0.053)	(0.034)	(0.061)	(0.125)	(0.059)	(0.084)
Within Province $\times$ Tier1	2.977***	2.867***	4.202***	1.632***	2.478***	3.331***
	(0.05)	(0.017)	(0.042)	(0.056)	(0.03)	(0.049)
Within Province $\times$ Tier2	3.085***	3.095***	3.917***	3.701***	3.946***	4.231***
	(0.036)	(0.014)	(0.031)	(0.064)	(0.03)	(0.045)
Hukou Index	-1.194***	-1.716***	-1.105***	-1.118***	-1.262***	-1.409***
	(0.004)	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)
Hukou Index^2	0.132***	0.236***	0.189***	0.139***	0.179***	0.262***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)

Back

Second Step: Lasso for each endogenous variable

Variables	Wage (Low-Skilled)	Wage (High-Skilled)	Rent	Amenity
Employment Shift-Share	✓		✓	
Migrant Shift-Share				✓
Export Shift-Share	✓			
Citation Shift-Share		✓		
Patent Shift-Share	✓			✓
Patent IV		✓	✓	✓
Import Shift-Share	✓		✓	✓
Net Export Shift-Share		✓	✓	✓
Trade Volume Shift-Share	✓		✓	✓
Altitude			✓	✓
TRI	✓	✓		
Undevelopable Land	✓			
Land Supply	✓		✓	✓
Export Shift-Share × Land Supply	✓			✓
Trade Volume Shift-Share × Land Supply			✓	✓
Robot Shift-Share × TRI			✓	
Migrant Shift-Share × Robot Shift-Share			✓	
Land Supply × Altitude	✓	✓		
Land Supply × TRI	✓			✓
Land Supply × Undevelopable Land	✓	✓	✓	✓
Wage Shift-Share × Altitude			✓	
Wage Shift-Share × Undevelopable Land				✓
Patent Shift-Share × Slope	✓			
Patent Shift-Share × Undevelopable Land				✓
Export Shift-Share × Slope		✓		
Export Shift-Share × TRI				✓

Back

Second Step: Lasso for each endogenous variable

VARIABLES	(1) First Stage R-squared	(2) First Stage F-statistics
Δ Log(High-skilled Wage)	0.677	159.77
Δ Log(Low-skilled Wage)	0.539	45.07
Δ Log(Housing Rent)	0.301	19.82
Δ Amenity Index	0.189	9.92

Back

Table 14: Estimation of Workers' Location Choice (Second Stage)

VARIABLES	(1) $\Delta\delta_{high}$	(2) $\Delta\delta_{low}$	(3) $\Delta\delta_{high}$	(4) $\Delta\delta_{low}$
Δ Log(High-skilled Wage)	0.328**		1.161**	
	(0.157)		[0.583]	
Δ Log(Low-skilled Wage)		0.322		1.673**
		(0.227)		[0.700]
Δ Log(Housing Rent)	0.148	0.165	-0.929*	-0.954*
	(0.101)	(0.128)	[0.511]	[0.504]
Δ Amenity Index	0.0656	0.0733	0.371**	0.196
	(0.0524)	(0.0615)	[0.164]	[0.266]
Constant	0.447***	0.557***	0.471***	0.358
	(0.125)	(0.167)	[0.179]	[0.260]
Observations	476	476	451	451
R-squared	0.035	0.033	0.055	0.058
Model	OLS	OLS	IV	IV

Back

Table 15: Labor Supply Estimation (Sector Choice)

Skill	Coefficient on w^a
Low-Skilled	1.048***
	(0.005)
High-Skilled	1.016***
	(0.013)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Back

Model Fit

Table 16: Model Fit

Variables	Model	Data	Difference
Panel A. Migration across Cities			
Total Migrants	26644291	26738972	-0.35%
Total High-skill Migrants	5746082	5830542	-1.45%
Total Low-skill Migrants	20898208	20908432	-0.049%
Panel B. Average Wages			
Mean Wages of High-skill in Agr	12651	12664	-0.11%
Mean Wages of High-skill in Non-agr	57090	56844	-0.43%
Mean Wages of Low-skill in Agr	12651	12664	-0.11%
Mean Wages of Low-skill in Non-agr	47076	46812	-0.57%
Panel C. Average Housing Rent and Amenity			
Mean Housing Rent	3343.6	3340.9	-0.08%
Mean Amenity	1.356	1.351	0.38%

Back

Model Fit

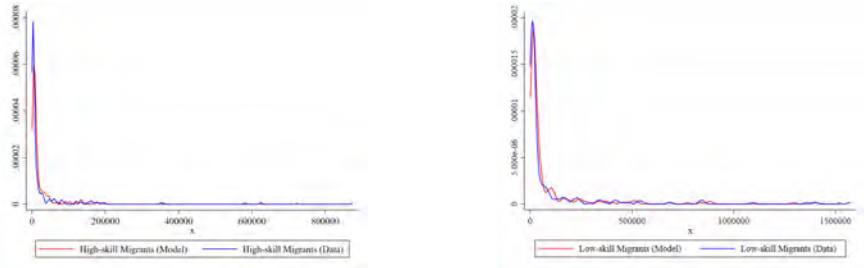


Figure 18: Model Fit of Migrants

Back

Model Fit

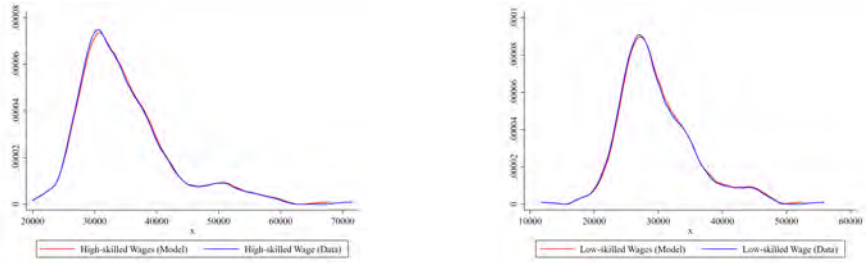


Figure 19: Model Fit of Wages

Back

Model Fit

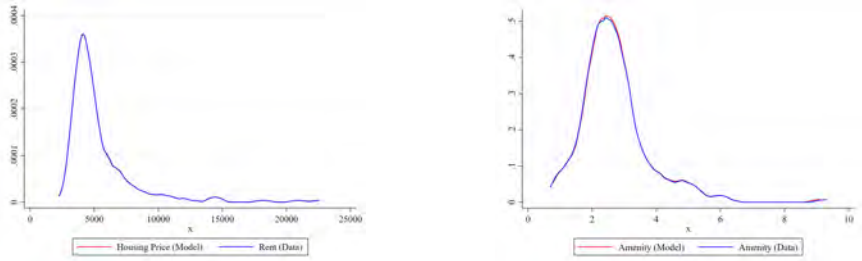


Figure 20: Model Fit of Housing Price and Amenity

Back