Quantitative Spatial Economics III: Diamond Style Model in Urban Economics

Zibin Huang [1]

[1]College of Business, Shanghai University of Finance and Economics

April 12, 2026

Overview

Introduction

- In the next two weeks, we are going to introduce Diamond style model
- We have two learning targets for this
- The first is to understand how to apply QSEM to traditional topics in urban economics
- The second is to understand how to transform a complicated structural model to a simple one

Introduction

- Sometimes we do not care that much about the primitive/deep parameters
- Meanwhile we want to implement policy counterfactual, what can we do?
- The basic idea is to transform a non-linear model to a linear one
- Then we can simply use reduced-form approach to estimate parameters one by one
- It is a good starting example for beginners to learn structural model

Introduction

- The original model comes from Diamond (2016), who investigates how spatial sorting exacerbated inequality in the U.S.
- She finds that during the last few decades, high-skilled and low-skilled workers sorted into different locations in America

Introduction

- High-skilled workers concentrated more and more in developed cities, increasing amenities and living costs
- Low-skilled workers were kicked out and had to live in cities with low amenities and living costs
- She shows that the overall inequality change was much larger than purely education wage gaps
- The increased inequality in amenity amplified the overall inequality

Motivation

- China experienced high economic growth during the last forty years
- TFP and technology developed rapidly
- Did this growth lead to similar spatial sorting and gentrification?
- Now we will introduce an application of Diamond model in China's context

Research Question

We have the following main research question:

- What is the impact of the technology shocks on migration for different skilled groups in China

We answer this question by constructing a spatial GE model with:

- Endogenous labor market
- Endogenous housing market
- Endogenous 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 much more low-skilled migrants than high-skilled migrants.
- Structural estimation results
  - Low-skilled workers care more about wages and housing prices, while high-skilled workers care more about amenities.
  - A positive shock in patents attracts more low-skilled workers than high-skilled workers, reduces the skill ratio and amenities, and discourages high-skilled migrants.

Preview of Findings

- Counterfactual of reducing patent from 2015 level to 2005 level:
  - Large reduction in low-skilled migration but not high-skilled migration
  - Welfare loss for both skills, especially low-skilled people with non-ag hukou
- Generally, growth in China during 2005 to 2015 is inclusive without diversification

Data

- Data Source:
  - Migration: micro-level Census 2005, 2010, 2015
  - Wages, housing prices, and amenities: statistical yearbooks
  - Patent citation data: China National Intellectual Property Administration and Google Patent
- Migrants: those who left hukou city for at least 6 months.
- Laborforce: age 25 - 50, currently working
- Non-Agriculture sector: working in urban regions and non-agriculture industries
- Agriculture sector: working in rural regions

PCA Results of the Amenity Index

	Loading	Unexplained variance
Panel A: Healthcare Index		
Hospital per 10,000 residents	0.7071	0.4351
Doctors per 10,000 residents	0.7071	0.4351
Panel B: Infrastructure Index		
Kilometers of road per 10,000 residents	0.4178	0.8078
Highway passengers per 10,000 residents	0.5987	0.6053
High-speed railway	0.6834	0.4856
Panel C: Environment Index		
PM 2.5	0.5315	0.339
Heavily polluted days	0.5712	0.2365
Polluted days	0.6255	0.08434
Panel D: Education Index		
Teacher-student ratio in primary schools	0.0818	0.9824
Teacher-student ratio in middle schools	0.1136	0.966
Number of colleges	0.5395	0.2333
Number of Project 985 universities	0.5886	0.08752
Number of Project 211 universities	0.5855	0.09703
Panel E: Amenity Index		
Healthcare Index	0.6434	0.4391
Infrastructure Index	0.5535	0.5848
Environment Index	-0.2340	0.9258
Education Index	0.4742	0.6952

Descriptive Analysis: Spatial Distribution of Patent Shock

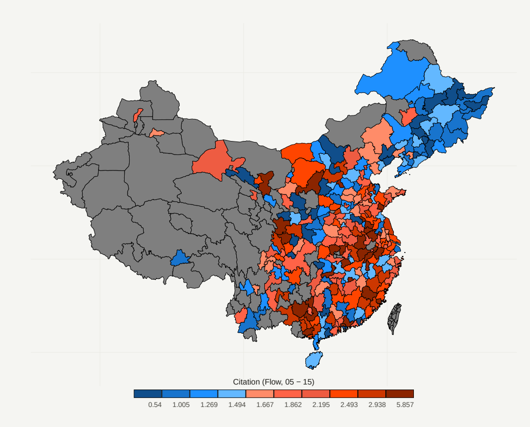

Figure 1: Spatial Distribution of Δ Log(Citation) (2005 - 2015)

▸ Patents

Descriptive Analysis

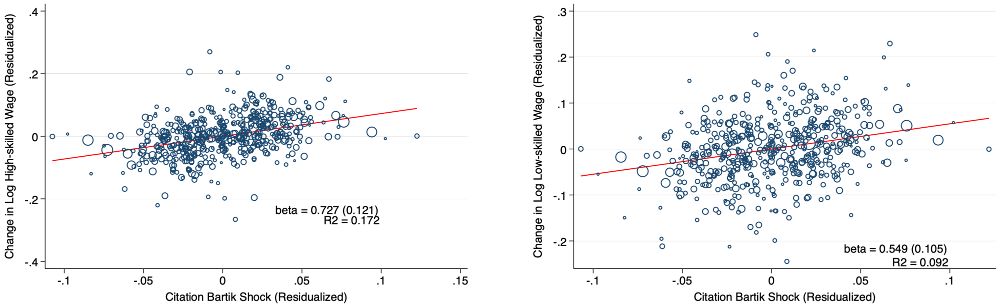

Figure 2: Effect of Citation Shock on Wages for High- and Low-skilled Workers

▸ Patents

Descriptive Analysis

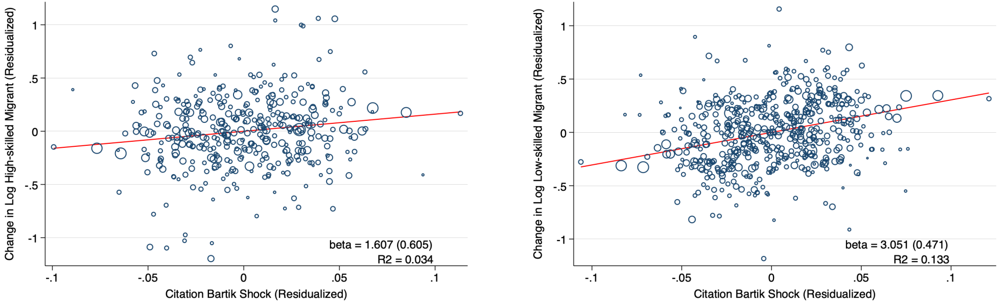

Figure 3: Effect of Citation Shock on Number of High- and Low-skilled Migrants

▸ Patents

Descriptive Analysis

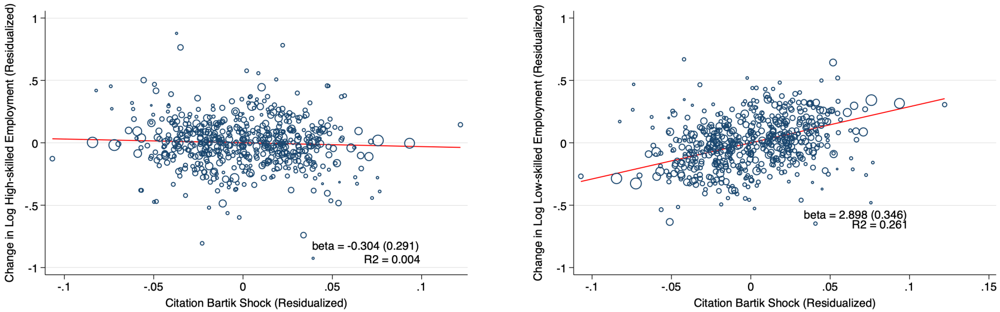

Figure 4: Effect of Citation Shock on Number of High- and Low-skilled Employment

▸ Patents

Descriptive Analysis

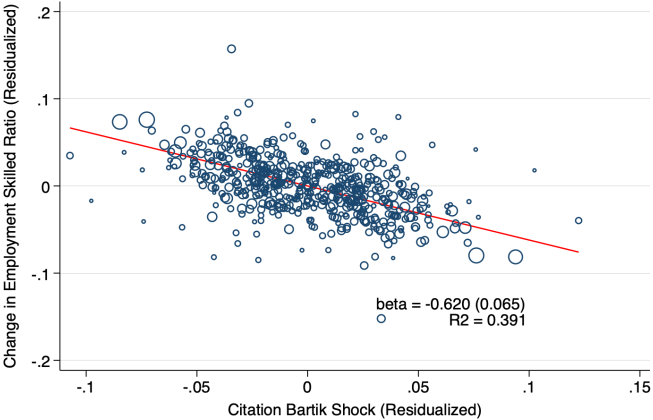

Figure 5: Citation Shock and Change in Skilled Ratio

Descriptive Analysis

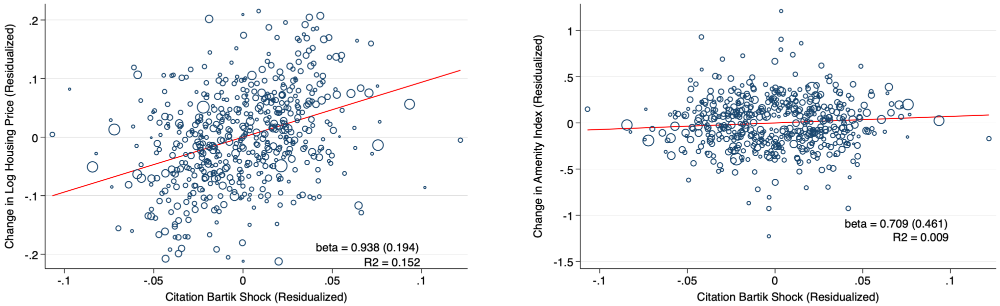

Figure 6: Effect of Citation Shock on Housing Price and Amenity

▸ Patents

Descriptive Analysis: Regression Analysis

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(Citation)	0.0477	-0.00261	0.0644	0.0545	0.0921	-0.0134*
	(0.0325)	(0.0313)	(0.0394)	(0.0757)	(0.0570)	(0.00785)
Panel B: Reduced Form						
Citation shock	1.837***	-0.304	2.898***	1.607***	3.051***	-0.620***
	(0.312)	(0.291)	(0.346)	(0.605)	(0.472)	(0.0654)
Panel C: IV						
Δ Log(Citation)	1.739**	-0.369	2.836**	1.056*	3.044**	-0.626**
	(0.772)	(0.357)	(1.240)	(0.565)	(1.441)	(0.276)
Year FE	X	X	X	X	X	X
City FE	X	X	X	X	X	X

Descriptive Analysis: Regression Analysis

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(Citation)	-0.00957	-0.0140	-0.0239	0.0889
	(0.0158)	(0.0167)	(0.0160)	(0.0630)
Panel B: Reduced Form				
Citation shock	0.727***	0.549***	0.938***	0.709
	(0.121)	(0.105)	(0.194)	(0.462)
Panel C: IV				
Δ Log(Citation)	0.727**	0.560*	1.017	0.955
	(0.352)	(0.292)	(0.623)	(0.688)
Year FE	X	X	X	X
City FE	X	X	X	X

Descriptive Statistics

- We have the following findings in our data
- As technology grew in China during the last decade:
  - Wages for both skills increased, low-skilled workers migrated more
  - Housing price increased
  - Amenity increased, but not in a very significant magnitude
- In general, we do not find positive sorting of migration resulting from the technology growth in China during 2005-2015
- This is totally different from previous findings in developed countries
- Why is this the case?
- Let's go to a structural model to explain these findings

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

Labor Demand — Non-agricultural Sector

- Firms in the non-agricultural sector 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 $C_{na,kt}$

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

- $N_{na,kt}^{\alpha_{na}}$ is a CES aggregator of labor
- L and C are substitutes
- θ is factor-augmenting productivity

Labor Demand — Non-agricultural Sector

- This is a non-linear CES+CD style production function with many primitive parameters
- In previous lectures, we directly estimate/calibrate these primitive parameters
- However, we can also take a detour if we only need the relation between labor force, technology, and output
- But not these deep parameters

Labor Demand — Non-agricultural Sector

- We can transform this production function to be a linear labor demand equation
- Using first order conditions and log linearization
- By F.O.C. we have:

$$
\begin{aligned}
W^H_{na,kt} &= z_{na,kt}\alpha N^{\alpha-\rho}_{na,kt}(\theta^K_{kt}K_{kt})^{1-\alpha}H^{\rho-1}_{na,kt}\theta^H_{kt} \\
W^L_{na,kt} &= z_{na,kt}\alpha N^{\alpha-\rho}_{na,kt}(\theta^K_{kt}K_{kt})^{1-\alpha}(L_{na,kt}+\omega C_{kt})^{\rho-1}\theta^L_{kt} \\
\kappa_t &= z_{kt}N^{\alpha}_{na,kt}(\theta^K_{kt}K_{kt})^{-\alpha}(1-\alpha)\theta^K_{kt}
\end{aligned}
$$

Labor Demand — Non-agricultural Sector

- We then log linearize this system and have labor demand functions as:

$$
\begin{aligned}
w^H_{na,kt} &= \ln W^H_{na,kt} = d_{na,kt} + (1-\rho)\ln N_{na,kt} + (\rho - 1)\ln H_{na,kt} + \ln \theta^H_{kt} \\
w^L_{na,kt} &= \ln W^L_{na,kt} = d_{na,kt} + (1-\rho)\ln N_{na,kt} + (\rho - 1)\ln(L_{na,kt} + \omega C_{kt}) + \ln \theta^L_{kt} \\
N_{na,kt} &= (\theta^L_{kt}(L_{na,kt} + \omega C_{kt})^{\rho} + \theta^H_{kt} H^{\rho}_{na,kt})^{\frac{1}{\rho}} \\
d_{na,kt} &= \ln \left(z^{1/\alpha}_{na,kt}\alpha \left(\frac{(1-\alpha)\theta^K_{kt}}{\kappa_t} \right)^{\frac{1-\alpha}{\alpha}} \right)
\end{aligned}
$$

Labor Demand — Non-agricultural Sector

- We can write the wage as a additive separable function of workforces H, L and technology A:

$$w^H_{na,kt} = g_{na,H}(A_{kt}, H_{na,kt}, L_{na,kt}) + d^H_{na,kt}$$
$$w^L_{na,kt} = g_{na,L}(A_{kt}, H_{na,kt}, L_{na,kt}) + d^L_{na,kt}$$

- d is the structural residual, determined by various of deep parameters in the production function
- Then we can use a simple log linear function to approximate them

Labor Demand — Non-agricultural Sector

- We further rewrite it a little bit:

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

- These are just two linear regressions!
- Log wage is y, log workforce numbers are x, error term d
- A_{kt} is proxied by the patent shock

Labor Demand — Non-agricultural Sector

- If we do not care so much about the details of the production function
- If you just want to have the relation between wage and labor demand at equilibrium
- We can just estimate regression coefficents γ using IV
- Instead of estimating the primitive parameters in the production function θ, z, ω
- This makes your life much much much easier

Labor Demand — Non-agricultural Sector

- But this is not feasible when you want to look deep into the production process
- Or if you want to run some counterfactuals by changing z or other deep parameters
- Therefore, this log-linearization trick can only help you to answer specific questions not related to deep parameters
- Always remember the tradeoff! If you want to answer more questions, the method can be more complicated
- Here in the Diamond model, we take a middle ground between reduced-form and fully structural approach

Labor Demand - Agricultural Sector

- The production in the agricultural sector only involves high-skilled labor $H_{a,kt}$ and low-skilled labor $L_{a,kt}$
- H and L are perfect substitutes

$$Y_{a,kt} = z_{ag,kt}(N_{a,kt}^{\alpha_a})^{\eta}$$
$$N_{a,kt} = L_{a,kt} + H_{a,kt}$$

- Similarly, we can transform labor demand function in agricultural sector as:

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

Labor Supply

- Four types of workers: ag local, ag migrant, na local, na migrant
- Two restrictions in location choice set
  - Workers with non-ag hukou do not choose ag sector
  - Workers with ag hukou do not choose ag sector in other locations
- Thus, workers with non-ag hukou choose among non-ag sectors in all locations in one step
- Workers with ag hukou make sequential decisions
  - Step 1: Choose between hometown ag sector and non-ag sector
  - Step 2: If choosing non-ag sector, which location to go

Labor Supply: Location Choices

The utility of working in the non-agricultural sector in city k and year t is

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

- w_{kt}^e is wage, r_{kt} is housing price; ζ is expenditure share on housing
- a_{kt} and ν_{kt} are the observed endogenous/unobserved exogenous amenity
- ϵ_{ikt} is an i.i.d. shock with T1EV distribution
- Heterogeneous preference varies by worker's skill e

Labor Supply: Location Choices

Migration cost can be decomposed into:

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

- $WithinHometown_{ikt}$ is the indicator to stay in home city
- $WithinProvince_{ikt}$ is the indicator to stay in home province
- The coefficients for them are different across skills and home regions
- $hukou_{ikt}$ is hukou policy in destination location

Labor Supply: Location Choices

- The value of working in the non-agricultural sector for individual i in year t with hometown k_0 is

$$W^{na}_{ik_0t} = \max\{V^{k_0}_{i1t}, V^{k_0}_{i2t}, ...V^{k_0}_{iKt}\}$$

  which is the maximum value of working in the non-agricultural over all possible cities.

- Based on the property of T1EV distribution, we have:

$$E[W^{na}_{ik_0t}] = \max\{V^{k_0}_{i1t}, V^{k_0}_{i2t}, ...V^{k_0}_{iKt}\} = ln[\sum_{k \in K} exp(V^{k_0}_{ikt})]$$

Labor Supply: Sector Choices

- For ag hukou workers, they have an additional first step sector choice
- The value of working in the agricultural sector in prefecture k is

$$W_{it}^{a} = \alpha_{0k_0} + \alpha_1 w_{a,k_0 t}^{e} + \xi_{it}^{a}$$
$$W_{it}^{na} = E[W_{ik_0 t}^{na}] + \xi_{it}^{na}$$

- α_{0k} is a city-specific constant term
- $w_{a,kt}$ is the agricultural earnings
- ξ_{it} is an i.i.d. shock with T1EV distribution
- The sector choice decision is

$$\max\{W_{it}^{na}, W_{it}^{a}\}$$

Labor Supply: Sector Choices

- This is a very typical discrete choice model
- It is non-linear, but still we have many off-the-shelf tools to estimate it
- We will show later how to estimate this non-linear system using IV: BLP method

Housing Supply

- Developers are price-takers and sell homogeneous houses at the marginal cost of production

$$R_{kt} = \iota_t \times MC(CC_{kt}, LC_{kt})$$

where ι_t is interest rate, CC_{kt} is construction costs, and LC_{kt} is land costs.

- The cost of land LC_{kt} is a function of the aggregate demand for local goods

$$HD_{kt} = L_{na,kt} W^{L}_{na,kt} + H_{na,kt} W^{H}_{na,kt}$$

- The housing supply equation is

$$r_{kt} = \ln(R_{kt}) = \ln(\iota_t) + \ln(CC_{kt}) + \gamma_k \ln(HD_{kt})$$
$$\gamma_k = \gamma_1^{hd} + \gamma_2^{hd} geo_k$$

where x_k^{geo} is the altitude that affects the elasticity of housing price with respect to local good demand.

Amenity Supply

- Endogenous amenity depends on technology and skill ratio:

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

- Local amenities respond to the education of neighboring households (Bayer, Ferreira, and McMillan, 2007) and college employment ratios (Diamond, 2016).
- Patents/tech growth have a direct impact on amenities.

Equilibrium Definition

Equilibrium in this model is defined by a set of working populations, wages, housing prices, and amenities such that

- The high-skill labor demand equals high-skill labor supply for both sectors and all cities.
- The low-skill labor demand equals low-skill labor supply for both sectors and all cities.
- Housing demand equals housing supply in the non-agricultural sector for all cities.
- Endogenous amenities demand equals endogenous amenity supply for both sectors and all cities.

Estimation Results

- We log linearize and take the first difference of these equations
- Then use different Bartik IVs to estimate the parameters of the model
- In estimation, we find that:
  - Patent growth increases wages ▸ Estimation of Labor Demand
  - Patent growth increases housing prices ▸ Estimation of Housing Supply
  - Patent growth increases amenity ▸ Estimation of Amenity Supply
  - Low-skilled care more about wage; High-skilled care more about amenity ▸ Estimation of Labor Supply
- The model fit is good ▸ Model Fit

Estimation of Labor Demand in the Non-agricultural Sector

- We use first difference regression to estimate labor demand equations

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

- We cannot use OLS here because structural residual $\Delta\epsilon$ is endogenous (function of A, H, L)
- Need variation in labor supply uncorrelated with unobserved changes in local productivity
- Instrument for low- and high-skilled workforce: migrant Bartik
- Instrument for local patent shock: patent Bartik

Back

Migrant Bartik

- 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.

▸ Back

Patent Bartik 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

Estimation of Labor Demand in the Non-agricultural Sector

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 City Population	(4) Δ Log Low-skilled City Population	(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-skilled Workers			0.627***	0.409***	0.344		
			(0.143)	(0.113)	(0.250)		
Migrant Bartik for Low-skilled Workers			-0.228	0.614***	-0.182		
			(0.255)	(0.201)	(0.446)		
Observations	468	468	468	468	468	468	468
Model	OLS	OLS	First stage	First stage	First stage	IV GMM	IV GMM
Sanderson-Windmeijer F			15.32	16.28	11.22		

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

Back

Estimation of Labor Demand in the Agricultural Sector

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

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

Back

Estimation of Labor Demand in the Agricultural Sector

Table 2: Estimation of the Labor Demand in the Agricultural Sector

VARIABLES	(1) Δ Log Agr Income	(2) Δ Log Agr Employ-ment	(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	481	481	481
R-squared	0.007		
Model	OLS	First stage	IV GMM
Sanderson-Windmeijer F		488.9	

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

Back

Estimation of the Housing Market

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

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

Back

Wage Bartik

- We interact cross-sectional differences in industrial employment composition with national changes in industry wage levels to construct the wage Bartik 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

Back

Estimation of the Housing Market

Table 3: Estimation of the Housing Market

VARIABLES	(1) Δ Log(Housing Price)	(2) Δ Log Housing Demand	(3) Δ Log Housing Demand * Geo	(4) Δ Log(Housing Price)
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 Bartik IV for High-skilled Workers		-5.737***	-18.66***	
		(1.235)	(6.272)	
Wage Bartik IV for Low-skilled Workers		6.693***	17.16***	
		(1.226)	(6.228)	
Wage Bartik IV for High-skilled Workers × Log Altitude		0.956***	3.284***	
		(0.211)	(1.069)	
Wage Bartik IV for Low-skilled Workers × Log Altitude		-0.948***	-2.046*	
		(0.210)	(1.065)	
Observations	468	468	468	468
Model	OLS	First stage	First stage	IV GMM
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

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

- Instrument for changes in the skilled ratio: employment Bartik for low- and high-skilled workers
- Instrument for patent shocks: patent Bartik

Back

Employment Bartik

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

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

- $H_{ind,na,-k,t}$ 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.

Back

Estimation of the Amenity Market

Table 4: Estimation of the Amenity Market

VARIABLES	(1) Δ Amenity Index	(2) Δ Log High-skilled City Population 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)	
Wage Bartik IV for High-skilled Workers		0.077	-1.108	
		(0.059)	(0.678)	
Wage Bartik 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
Sanderson-Windmeijer F		14.28	8.193	

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

Back

Labor Supply: Location Choices

- We have the following utility function to estimate:

$$
\begin{aligned}
V_{ikt} = {} & \beta_e^w (w_{na,kt}^e - \zeta r_{kt}) + \beta_e^h WithinHometown_{ikt} + \beta_e^p WithinProvince_{ikt} \\
& + \beta_e^h hukou_{kt} + \beta_e^a a_{kt} + \nu_{kt}^e + \epsilon_{ikt}
\end{aligned}
\tag{1}
$$

- With T1EV distribution, it is a typical Logit model
- We can always use MLE to estimate it
- However, we all know that wage, rent, amenity are all endogenous in the model
- They are correlated with unobserved amenity ν

Back

Endogeneity in DCM

- We must find a way to solve this endogeneity issue
- In linear regression, we usually use IV
- Therefore, we have to find a way to apply IV in discrete choice model
- Traditional linear IV method did not work in this case
- We introduce two methods today: BLP and Control Function

Back

Endogeneity in DCM: 1. BLP

- The first method is called BLP, introduced in Berry, Levinsohn, and Pakes (1995)
- It first transform the endogeneity issue in a nonlinear model to a linear one
- Then we can use well-developed linear IV method to solve it

Endogeneity in DCM: 1. BLP

- Assume we have the car buying problem
- There are M markets with J_m options (brands) in each market
- Utility for consumer n in market j to choose brand m is:

$$U_{njm} = V(p_{jm}, x_{jm}, s_n, \beta_n) + \xi_{jm} + \epsilon_{njm}$$

- p_{jm} price; s_n personal attributes; x_{jm} product attributes; ξ_{jm} unobserved product attributes; ϵ_{njm} i.i.d. T1EV shock
- Price is correlated with unobserved brand attributes $\xi_{jm} \not\perp p_{mj}$
- Usually we assume a linear functional form for V

Endogeneity in DCM: 1. BLP

- Important feature of BLP: endogeneity comes from market-product level ξ_{jm}
- That is why we can use it in Diamond model
- Migration choice is individual level
- But endogeneity comes from destination location level

Endogeneity in DCM: 1. BLP

- The idea of BLP employs a two-step approach
- First, add in a product-market level FE, absorb ξ_{jm}
- Estimate the equation with fixed effect
- Second, open the box of product-market level FE, estimate the remaining parameters

Endogeneity in DCM: 1. BLP

- We can decompose the observed utility value into

$$V_{njm} = \underbrace{\bar{V}(p_{jm}, x_{jm}, \bar{\beta})}_{\text{varies only over product-market}} + \underbrace{\tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n)}_{\text{varies also over consumer}}$$

- Then we have the utility to be

$$U_{njm} = \underbrace{[\bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \xi_{jm}]}_{\text{a product-market level fixed effect}} + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \epsilon_{njm}$$

- We just combine all terms varying only at product-market level together

Endogeneity in DCM: 1. BLP

- We define product-market level FE as:

$$\delta_{jm} = \bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \xi_{jm} \tag{2}$$

$$U_{njm} = \delta_{jm} + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \epsilon_{njm} \tag{3}$$

- Equation (3) does not entail any endogeneity and V is linear
- Step 1: We run a Logit model with *jm* level FE to estimate parameters $\tilde{\beta}$ and fixed effect δ_{jm}
- Step 2: We get estimates of δ_{jm} in step 1, and run IV regression for equation (2) and get $\bar{\beta}$

Endogeneity in DCM: 1. BLP

The essence of BLP

- We cannot run IV regression directly in DCM
- We first pack all terms at the level where endogeneity happens into FE
- Then we estimate a DCM with these FEs
- We have estimated FEs, then unpack it and run linear IV regression
- Transform non-linear IV to be linear IV
- BLP tells you how to use an IV in a DCM, but only in a specific model structure with nested endogeneous variables.

Endogeneity in DCM: 2. Control Function

- BLP is not always feasible (error structure...)
- The algorithm of estimating BLP is sometimes complicated
- If the dimension of the fixed effect is too high, you have to use some contraction method
- Highly recommend you to read BLP part in Train's book (or better, BLP 1993)
- The second important non-linear IV approach is Control Function (CF)

Endogeneity in DCM: 2. Control Function

- The utility of consumer n buying product j is:

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \epsilon_{nj}$$

- y_{nj} is endogenous, $y_{nj} \not\perp \epsilon_{nj}$
- We assume that there is an instrument z_{nj}, related with y_{nj} by first stage:

$$y_{nj} = W(z_{nj}, \gamma) + \mu_{nj} \tag{4}$$

- Assume that $\epsilon_{nj}, \mu_{nj} \perp z_{nj}$, $\epsilon_{nj} \not\perp \mu_{nj}$
- $\epsilon_{nj} \not\perp \mu_{nj}$ implies that y_{nj} and ϵ_{nj} are correlated

Endogeneity in DCM: 2. Control Function

- Therefore, μ is the source of the endogeneity
- We want to to some extent control it
- We can do a CEF decomposition (given μ_{nj}) for ϵ_{nj}:

$$\epsilon_{nj} = \underbrace{E(\epsilon_{nj}|\mu_{nj})}_{CF(\mu_{nj},\lambda)} + \tilde{\epsilon}_{nj}$$

- By construction: $E[\tilde{\epsilon}_{nj}|\mu_{nj}] = 0$
- Thus, we have $\tilde{\epsilon}_{nj} \perp y_{nj}$ (y is correlated with ϵ only through μ)
- We call $CF(\mu_{nj}, \lambda)$ a control function, where λ is some parameter

Endogeneity in DCM: 2. Control Function

- Then we can rewrite the utility function as

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + CF(\mu_{nj}, \lambda) + \tilde{\epsilon}_{nj} \tag{5}$$

- Step 1: Estimate first stage equation (4), get residual of the first stage $\hat{\mu}$
- Step 2: Plug $\hat{\mu}$ in the CF (5)
- Step 3: Estimate equation (5) using simple Logit
- In step 2, we need to assume a functional form for CF
- Usually we can choose flexible non-parametric form (e.g. high-order polynomials)

Endogeneity in DCM: 2. Control Function

- The logic of CF approach is as follows:
  - We know that instrument z is not correlated with the error ϵ
  - Thus, endogenous variable y correlates with ϵ only through first stage error μ, but not IV z
  - Then by controlling the correlated parts of μ and ϵ, we can eliminate the correlation of y and ϵ
- CF is a pretty general method
- But it requires you to set a function form for CF

Labor Supply: Location Choices

- Now we go back to our paper
- We choose BLP method to tackle the endogeneity issue
- Our utility function has a nested error structure
- Choice is made at individual level
- Endogeneity happens at location level: wages, housing prices, amenities, and Hukou policies
- So we can estimate the model by adding location fixed effect, then regress this fixed effect on endogeneous variables with IV

Labor Supply: Location Choices

BLP model:

$$V_{ikt}^{k_0} = \beta_h^e WithinHometown_{ikt} + \beta_p^e WithinProvince_{ikt} + \delta_{kt}^e + \epsilon_{ikt}$$
$$\delta_{kt}^e = \beta_w^e(w_{kt}^e - \zeta r_{kt}) + \beta_h^e \text{Hukou} + \beta_a^e a_{kt} + \nu_{kt}^e$$

- First step: maximum likelihood, get δ_{kt}^e.
- Second step:

$$\Delta\delta_{kt}^e = \beta_w^e(\Delta w_{kt}^e - \zeta \Delta r_{kt}) + \beta_h^e \Delta\text{Hukou} + \beta_a^e \Delta a_{kt} + \Delta\nu_{kt}^e$$

  - ζ: expenditure share on housing: mortgage share 40%, rent share 30%, set $\zeta = 0.35$; alternatively, set $\zeta = 0.62$ following Diamond (2016).
  - Instruments: choose from a large set of potential IVs using Lasso (Chernozhukov et al. 2018).

Back

Table 5: Labor Supply Estimation (Location Choice, BLP first stage)

Worker Type Year	Low-Skilled 2005	Low-Skilled 2010	Low-Skilled 2015	High-Skilled 2005	High-Skilled 2010	High-Skilled 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

Table 6: First stage statistics of BLP second stage IV regression

VARIABLES	(1) First Stage R-squared	(2) First Stage F-value
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

Labor Supply: Location Choices

Table 7: Labor Supply Estimation (Location Choice, BLP 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]
Observations	476	476	451	451
R-squared	0.035	0.033	0.055	0.058
Model	OLS	OLS	IV	IV

Back

Labor Supply: Sector Choices

Table 8: Labor Supply Estimation (Sector Choice)

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

Back

Mechanism

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

Solving the Equilibrium: Algorithm

- The endogenous variables of city k in year t include $\mathbf{\Delta_0} = \{\mathbf{H_0}, \mathbf{L_0}, \mathbf{W_0}, \mathbf{r_0}, \mathbf{a_0}\}$
- Let $N_{k_0}^{a,s}$ and $N_{k_0}^{na,s}$ be the number of skill s agricultural and non-agricultural hukou workers from hometown city k_0
- Let q denote the iteration time
- Within each iteration, we use $\hat{var}$ to denote the temporary updating result of some variable var
- At the beginning of the q-th iteration, we have $\mathbf{\Delta_{q-1}}$.

Solving the Equilibrium: Algorithm

- Step 1: update workers' utility values using endogenous variables derived from the last iteration $(q-1)$

$$\begin{aligned}\hat{\delta}_k^e &= \beta_1^e w_{na,k|q-1}^e + \beta_2^e r_{k|q-1} + \beta_3^e a_{k|q-1} + \nu_k^e \\ \hat{V}_{ik}^{k_0} &= MigrationCost_{ik|q-1} + \hat{\delta}_k^e \\ E[\hat{W}_{ik_0}^{na}] &= ln[\sum_{k \in K} exp(\hat{V}_{ik}^{k_0})]\end{aligned}$$

Solving the Equilibrium: Algorithm

- Step 2: update migration flows using the logit-form migration equations

$$\hat{H}_k^{na} = \sum_{i \in N^{na,H}} \frac{\exp(\hat{V}_{ik})}{\sum_r^K \exp(\hat{V}_{ir})} + \sum_{i \in N^{a,H}} \frac{exp(\hat{W}_{ik_0}^{na})}{exp(\hat{W}_i^a) + exp(\hat{W}_{ik_0}^{na})} \cdot \frac{\exp(\hat{V}_{ik})}{\sum_r^K \exp(\hat{V}_{ir})}$$

$$\hat{H}_k^{a} = \sum_{i \in N_k^{a,H}} \frac{exp(\hat{W}_i^a)}{exp(\hat{W}_i^a) + exp(\hat{W}_{ik}^{na})}$$

$$\hat{L}_{kt}^{na} = \sum_{i \in N^{na,L}} \frac{\exp(\hat{V}_{ik})}{\sum_r^K \exp(\hat{V}_{ir})} + \sum_{i \in N^{a,L}} \frac{exp(\hat{W}_{ik_0}^{na})}{exp(\hat{W}_i^a) + exp(\hat{W}_{ik_0}^{na})} \cdot \frac{\exp(\hat{V}_{ik})}{\sum_r^K \exp(\hat{V}_{ir})}$$

$$\hat{L}_{kt}^{a} = \sum_{i \in N_k^{a,L}} \frac{exp(\hat{W}_i^a)}{exp(\hat{W}_i^a) + exp(\hat{W}_{ik}^{na})}$$

Solving the Equilibrium: Algorithm

- Step 3: update wages in each city using the wage equilibrium equation

$$\hat{w}^H_{ag,k} = \hat{w}^L_{ag,k} = \beta^1_0 + \gamma_{ag} \ln(\hat{H}_{ag,k} + \hat{L}_{ag,k}) + \epsilon_1$$
$$\hat{w}^H_{na,k} = \beta^2_0 + \gamma_{HA} A_k + \gamma_{na,LH} \ln \hat{H}_{na,k} + \gamma_{na,LL} \ln \hat{L}_{na,k} + \epsilon_2$$
$$\hat{w}^L_{na,k} = \beta^3_0 + \gamma_{LA} A_k + \gamma_{na,LH} \ln \hat{H}_{na,k} + \gamma_{na,LL} \ln \hat{L}_{na,k} + \epsilon_3$$

Solving the Equilibrium: Algorithm

- Step 4: update the housing price in each city using housing equilibrium equation

$$\hat{r}_k = \beta_0^4 + \left[\gamma_1^{hd} + \gamma_2^{hd} \times \ln(Slope_k)\right] \ln(\hat{L}_k^{na} e^{\hat{w}_{na,k}^L} + \hat{H}_k^{na} e^{\hat{w}_{na,k}^H}) + \epsilon_4$$

- Step 5: update the amenity in each city using the amenity determination equation

$$\hat{a}_k = \beta_0^5 + \gamma_1^a A_k + \gamma_2^a \ln\left(\frac{\hat{H}_k^{na}}{\hat{L}_k^{na}}\right) + \epsilon_5$$

Solving the Equilibrium: Algorithm

- Having these predicted values of the endogenous variables, we use the following updating rule to get the values of all variables for the next iteration:

$$\boldsymbol{\Delta_q} = \zeta \boldsymbol{\Delta_{q-1}} + (1-\zeta)\boldsymbol{\hat{\Delta}_{q-1}} \tag{6}$$

- $0 < \zeta < 1$, iterate until convergence is achieved

Counterfactual 1: Patents Reduce to 2005 Level

- Change patent to the 2005 level
- Simulate labor supply, wages, housing price, and amenities in the general equilibrium framework

Counterfactual 1: Patents Reduce to 2005 Level

Table 9: Eliminating Innovation Growth: Patent Citation Change in Log Points

	Mean	Std Dev	Max	Min
National	-1.748	1.048	1.792	-5.857
Eastern Region	-1.864	0.699	-0.231	-3.734
Middle Region	-2.105	0.993	0.274	-5.857
Northeastern Region	-0.646	0.879	1.792	-1.861
Western Region	-1.747	1.213	1.705	-5.412

Counterfactual 1: Patents Reduce to 2005 Level

Table 10: Eliminating Innovation Growth: Counterfactual Results Summary 1

	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 Workforce 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%

- Migration reduces, especially for low-skill workers
- Fewer rural-to-urban migrations, higher urban skill ratio
- Lower wages for everyone

Counterfactual 1: Patents Reduce to 2005 Level

Table 11: Eliminating Innovation Growth: Counterfactual Results Summary 2

	Original Eq	Counterfactual	Change
Panel D. Housing Rent and Amenity			
National Average Housing Rent	3343.61	1128.69	-66.24%
National Average Amenity	2.712	1.124	-58.6%
Panel E. Average Welfare Changes			
Welfare of Low-skilled Agr			-1.76%
Welfare of Low-skilled Non-agr			-16.81%
Welfare of High-skilled Agr			-4.52%
Welfare of High-skilled Non-agr			-13.49%
Panel F. Inequality			
National Wage Gini Coefficient	0.430	0.253	-41.16%
National Welfare Gini Coefficient	0.0965	0.103	6.74%

- Lower housing rent and amenity
- Lower wages for everyone, especially people with non-agr hukou
- Lower wage inequality, but higher welfare inequality

Counterfactual 1: Patents Reduce to 2005 Level

- The results show that low-skilled workers would be more damaged if we eliminate the innovation in China, more than high-skilled workers
- They migrate less, earn less, cannot enjoy amenity in urban regions
- It is very interesting to see that eliminating innovation could reduce wage inequality, but enlarge welfare inequality!
- This is because rural people stop migrating to urban regions!
- Therefore, we can see that the technology growth in China really helps low-skilled workers

Counterfactual 2: Channel Decomposition

- There are three main channels for technology to affect migration
  - Wage effect attracts migration
  - Direct amenity effect attracts migration
  - Indirect amenity effect through changes of skill ratio
- We now try to decompose the overall effect of tech growth on migration into these three channels

Counterfactual 2: Channel Decomposition

Table 12: Channel Analysis: Regression Coefficient Summary

	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)

Conclusion

- Patent growth 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
- Patent growth 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 gentrification and benefited workers with different skills

Summary Statistics

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Share of migrants among the working population	609	0.16	0.16	0.00	0.90
Share of migrants among the high-skilled working population	609	0.08	0.09	0.00	0.59
Share of migrants among the low-skilled working population	609	0.19	0.19	0.00	0.96
Citations of Patents	571	2903.20	11265.63	2.00	157306.00
Wages of workers in the agricultural sector (2015 CNY)	576	10072.41	4181.63	2638.21	26838.00
Wages of high-skilled workers in the NA sector (2015 CNY)	595	48127.27	14217.58	15928.67	122615.09
Wages of low-skilled workers in the NA sector (2015 CNY)	595	39474.81	11309.51	6007.17	91138.81
City-level average house price (2015 CNY)	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
With High-speed railway stations	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
Average uphill slope of terrain (%)	575	3.96	3.13	0.00	18.34

Initial Skilled Ratio and Change in Skilled Ratio

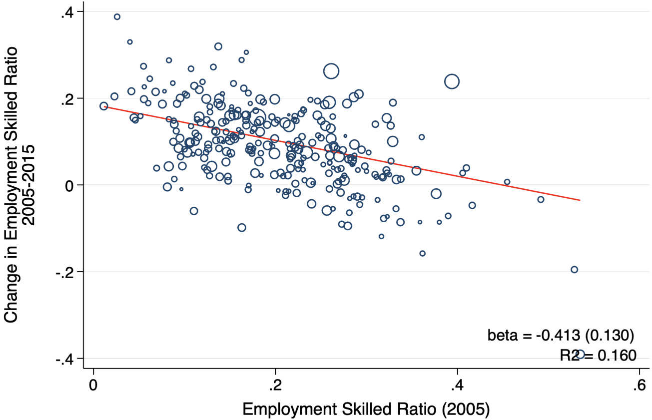

Figure 7: Initial Skilled Ratio and Change in Skilled Ratio

Model Fit

Table 13: 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

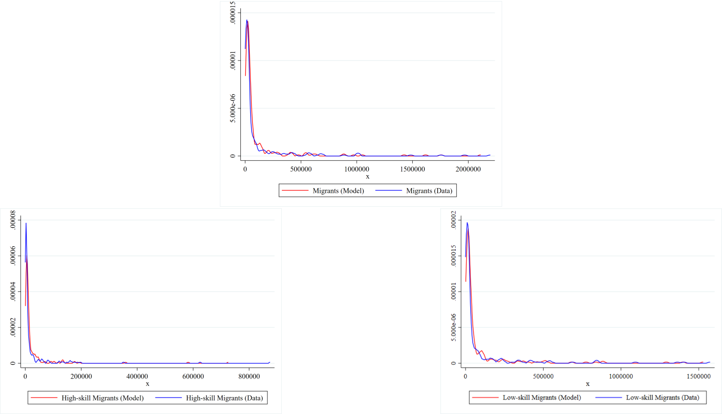

Figure 8: Model Fit of Migrants

Back

Model Fit

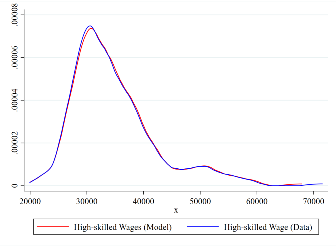

Figure 9: Model Fit of Wages

▸ Back

Model Fit

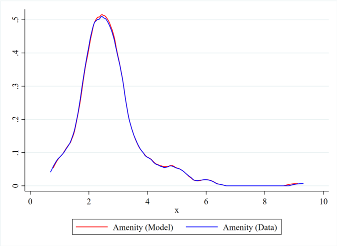

Figure 10: Model Fit of Housing Price and Amenity

▸ Back

Spatial Distribution of Patent Shock

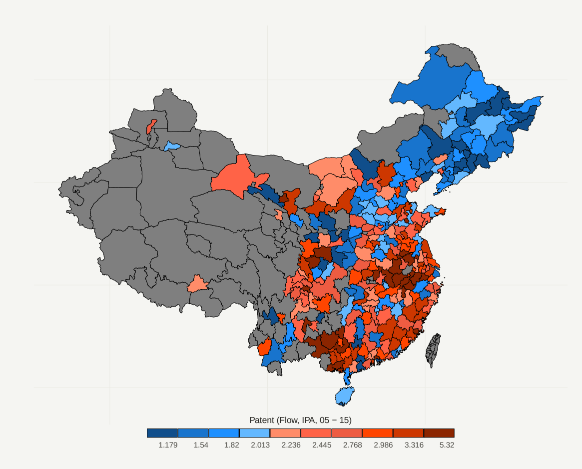

Figure 11: Spatial Distribution of Log Patent Growth (2005 - 2015)

Descriptive Analysis

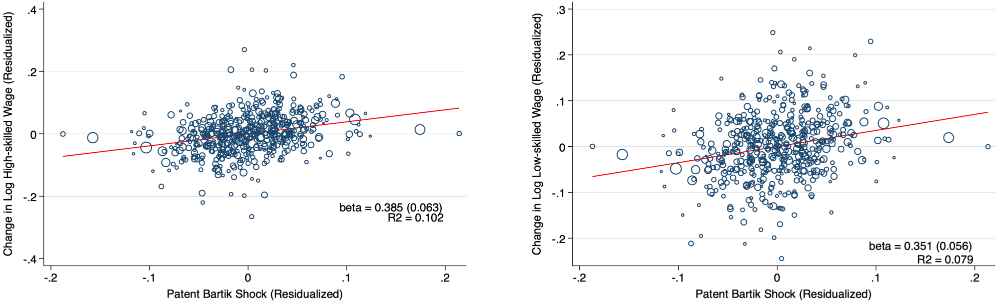

Figure 12: Effect of Patent Shock on Wages for High- and Low-skilled Workers

Back

Descriptive Analysis

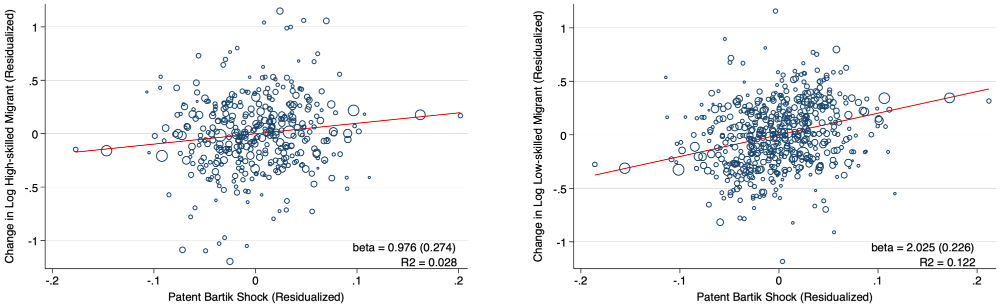

Figure 13: Effect of Patent Shock on Number of High- and Low-skilled Migrants

▸ Back

Descriptive Analysis

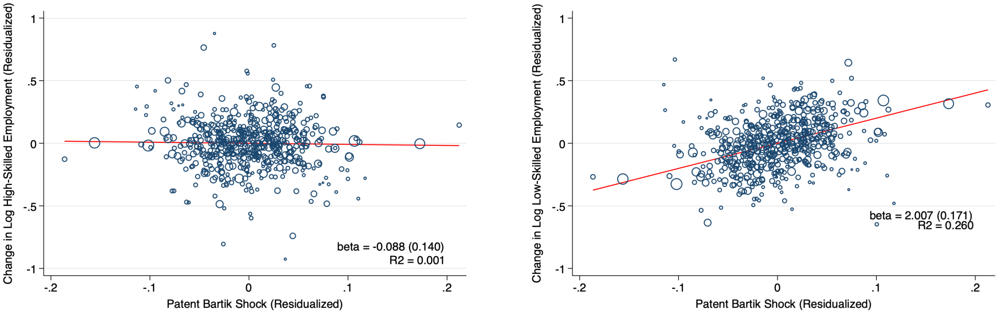

Figure 14: Effect of Patent Shock on Number of High- and Low-skilled Employment

Back

Descriptive Statistics

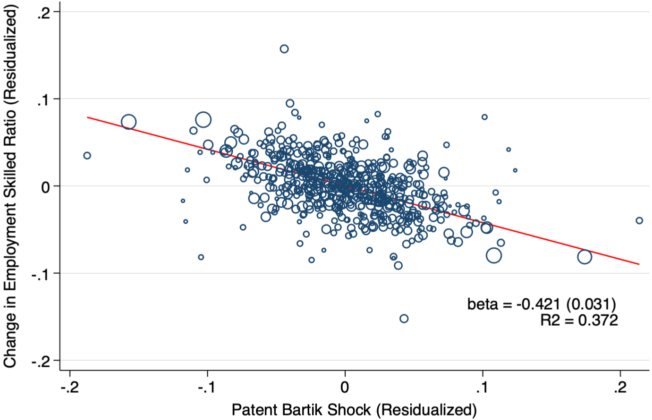

Figure 15: Patent Shock and Change in Skilled Ratio

Descriptive Analysis

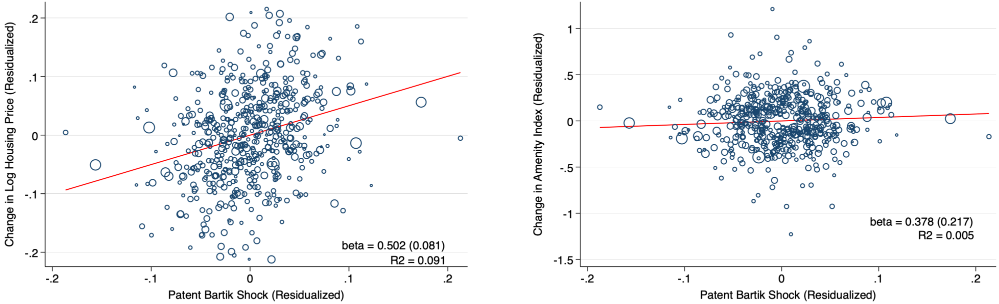

Figure 16: Effect of Patent Shock on Housing Price and Amenity

▸ Back

References

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115 (4):588–638.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4):841–890.

Card, David. 2009. "Immigration and Inequality." *American Economic Review* 99 (2):1–21.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *American Economic Review* 106 (3):479–524.