Size of Local Labor Market and Productivity: A Quantitative Analysis of China

Xiang Zhang

Thesis Advisor: Professor Richard Hornbeck

- Economic activity is spatially concentrated (Duranton and Puga[2004]) . Higher density of local economic activities creates productivity advantages (Glaeser and Gottlieb[2009], Greenstone, Hornbeck, and Moretti[2010])
- Larger labor markets generate agglomeration economy
  - The quality of matching between firms and workers is higher in larger labor markets (Diamond[1982], Helsley and Strange[1990])
  - Concentration of human capital generates positive externality and increases productivity (Lucas[1988], Glaeser[1999], Moretti[2004])

Introduction

- Larger labor markets generate agglomeration economy
  - The quality of matching between firms and workers is higher in larger labor markets (Diamond[1982], Helsley and Strange[1990])
  - Concentration of human capital generates positive externality and increases productivity (Lucas[1988], Glaeser[1999], Moretti[2004])
- Migration changes size of local labor markets, but also changes the composition of workers
  - Number of migrant workers in China: 37 million in 1997, 145 million in 2009, and 245 million in 2017 (National Bureau of Statistics), the majority of whom are low-skill workers
- Question: Do expanded local labor market size increases or decreases productivity?

Conceptual Framework

- The production function of of city c in year t can be written in the following CES function

$$Y_{ct} = K_{ct}^{\alpha}[(A_{ct}^{H}H_{ct})^{\frac{\sigma-1}{\sigma}} + (A_{ct}^{L}L_{ct})^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}(1-\alpha)}$$

  where K_{ct} is the capital stock, H_{ct} and L_{ct} represent the number of high-skill and low-skill workers respectively, and A_{ct} is the total factor productivity

- Rewrite the production function as

$$Y_{ct} = K_{ct}^{\alpha}[A_{ct}\phi_{ct}(H_{ct} + L_{ct})]^{(1-\alpha)}$$

  where

$$\phi_{ct} = [(\beta_{ct}h_{ct})^{\frac{\sigma-1}{\sigma}} + ((1-\beta_{ct})(1-h_{ct}))^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}}$$

- $A_{ct}\phi_{ct}$ represents the local productivity
- In this setting, β_{ct} captures the relative productivity between high-skill and low-skill workers, h_{ct} is the fraction of high-skill workers (Peri[2010])

Conceptual Framework

- Improved matching quality and knowledge spillover imply that productivity is an increasing function of labor market size

$$\frac{\partial A_{ct}\phi_{ct}}{\partial N_{ct}} > 0$$

- A higher proportion of high-skill workers will raise the overall productivity

$$\frac{\partial A_{ct}\phi_{ct}}{\partial h_{ct}} > 0$$

Measurement of City Productivity

- Assume a Cobb-Doglas productivity function at firm-level, I estimate city-level TFP by running the following regression using firm data (Greenstone, Hoenbeck, and Moretti [2010], Hornbeck and Moretti[2019])

$$\log(Y_{it}) = \beta_1 \log(L_{it}) + \beta_2 \log(K_{it}) + \beta_3 \log(M_{it}) + \gamma_{ct} + \varepsilon_{it}$$

  where Y_{it} is the total output of firm i in year t, L_{it} is the number of employee, K_{it} is the value of capital stock, and M_{it} is the material inputs

- The estimated city-by-year fixed effect γ_{ct} represents the average TFP of all firms in city c in year t
- I also use the following methods to estimate city TFP to check the robustness
  - Olley and Pakes Method (Olley and Pakes[1996])
  - Cost-sharing Method
  - Levinsohn and Petrin Method (Levinsohn and Petrin[2003])

Methodology and Instrumental Variables

- To analyze the effects of labor market size on productivity, I estimate the following specification

$$\log(A_{ct}) - \log(A_{ct-N}) = \alpha(\log(L_{ct}) - \log(L_{ct-N})) + \gamma_p + \varepsilon_{ct}$$

  where A_{ct} is the productivity of city c in year t, L_{ct} is population size, γ_p is the region fixed effect, and ε_{ct} is error term

- Taking the first-difference could attenuate endogeneity issue, but we still have factors that might correlate with both changes in TFP and changes in labor market size

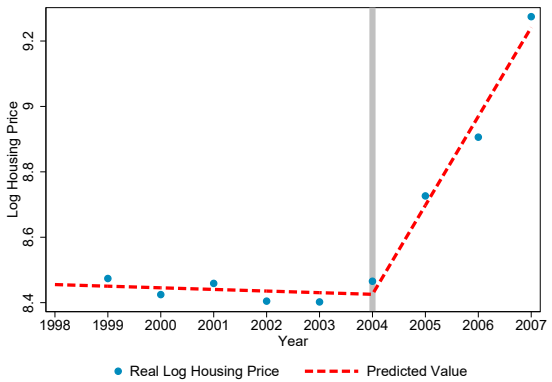
- First IV: structural break in housing price (Charles, Hurst, and Matthew[2018])

- Second IV: shift-share IV

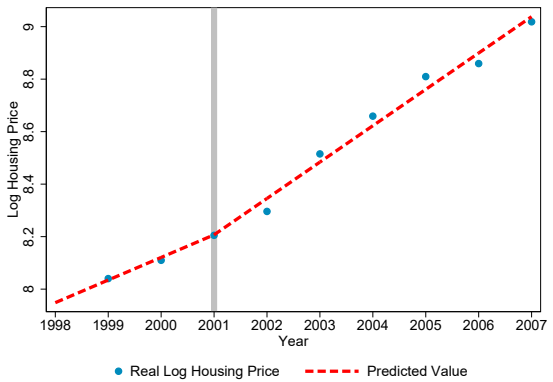
First IV: Structural Break in Housing Price

- Housing boom in China largely increases the housing price, but the magnitude varies (Chen and Wen[2018], Glaeser et al.[2017])
- Abnormal changes in housing price could be a result of speculative investment behavior, but not changes in foundamentals (Scheinkman and Xiong 2003, Brueckner et al. 2016)
- I estimate the following equation for each city, the λ_k represents the magnitude of structural break in housing price

$$P_k^H(t) = \alpha + \tau_k t + \lambda_k (t - t_k^*) \mathbb{1}\{t > t_k^*\} + \varepsilon_{kt}$$



(a) Beijing: Huge Jump



(b) Shanghai: Medium Jump

Second IV: Shift-share IV

- In shift-share IV, I instrument for population change during 2000 and 2010 using predicted population change (Altoji and Card[1991], Shimer[2001], Borjas[2003], Saiz[2007, 2010], Hornbeck and Moretti[2019])

$$IV_c = \alpha_{c,1990-2000} \times \Delta_{2000-2010}$$

where $\alpha_{c,1990-2000}$ is the share of net population change during 1990-2000 of city c, and $\Delta_{2000-2010}$ is the total net population change for all cities in China from 2000 to 2010

Data

- Firm Data: Annual Survey of Manufacturing Firms 1998-2007
  - Census of all manufacturing firms with more than 5 million Yuan sales value (600 thousand USD in 2000)
  - Sample size increased from 160 thousand in 1998 to 340 thousand in 2007
- City Population Data
  - Covers all 334 cities in China
  - Statistical yearbook of all provinces
  - 1990, 2000, and 2010 Census data bulletin
- City Housing Price Data
  - Covers 208 cities from 2000 to 2008 (some cities also have data for 1998 and 1999)
  - Statistical yearbook of all provinces

First IV Results: Effects of Labor Market Size on TFP

- Increase in population size lowers TFP
- The magnitude of negative effects gets smaller over time, suggesting labor market is responsive to influx of population

	(1) 2004-2007	(2) 2003-2007	(3) 2002-2007	(4) 2001-2007	(5) 2000-2007
Log Pop Difference: 2004-2007	-3.185***				
	(0.682)				
Log Pop Difference: 2003-2007		-1.876***			
		(0.631)			
Log Pop Difference: 2002-2007			-2.009***		
			(0.697)		
Log Pop Difference: 2001-2007				-1.637***	
				(0.542)	
Log Pop Difference: 2000-2007					-1.587***
					(0.552)
Observations	208	208	208	206	207
F	3.784	3.801	3.819	3.803	3.867
Prob > F	0.0007	0.0007	0.0007	0.0007	0.0006

Standard errors in parentheses

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