ESSAYS IN URBAN ECONOMICS: AGGLOMERATION ECONOMIES IN THE MANUFACTURING SECTOR AND LAND CONSTRAINTS IN HOUSING MARKETS

A Dissertation Presented to

the Faculty of the Department of Economics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Yilin Dong

May 2017

ESSAYS IN URBAN ECONOMICS: AGGLOMERATION ECONOMIES IN THE MANUFACTURING SECTOR AND LAND CONSTRAINTS IN HOUSING MARKETS

Yilin Dong

APPROVED:

Janet E. Kohlhase, Ph.D.
Committee Chair

Bent E. Sørensen, Ph.D.

Vikram Maheshri, Ph.D.

Justin H. Kirkland, Ph.D.
Department of Political Science
University of Houston

Antonio D. Tillis, Ph.D.
Dean, College of Liberal Arts and Social Sciences
Department of Hispanic Studies

ACKNOWLEDGEMENTS

My special thanks go to my advisor, Dr. Janet Kohlhase, for the continuous support of my Ph.D study and related research, for her immense knowledge, patience and motivation. I also owe my sincere appreciation to my committee members, Dr. Bent E. Sørensen and Dr. Vikram Maheshri for their advice and support throughout my Ph.D. study. I would like to thank Dr. Kirkland for being part of my committee as an outside reader, and providing insight comments.

I would also like to express my gratitude to my colleagues Chon-kit Ao who gave me many valuable suggestions. In the end, my deepest gratitude goes to my husband and my parents. Without their encouragement, love, and continuous support, I would not have finished this dissertation work.

ESSAYS IN URBAN ECONOMICS: AGGLOMERATION ECONOMIES IN THE MANUFACTURING SECTOR AND LAND CONSTRAINTS IN HOUSING MARKETS

An Abstract of a Thesis
Presented to

the Faculty of the Department of Economics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Yilin Dong

May 2017

ABSTRACT

This dissertation focuses on the study of urban development. The first chapter explores whether agglomerative forces can explain the location decisions of new manufacturing firms in the face of declining manufacturing activity in the United States over the time period 2004-2011. I find that labor market pooling and input-output linkages have the largest effects, positively influencing firm location. Moreover, corporate taxes discourage firm activity but the effects are weaker in more geographically concentrated industries. I then investigate whether negative macro shocks would change how firm location decisions respond to agglomeration forces. The results indicate that the workings of agglomeration economies have become more pronounced after the Great Recession. New firms may become more risk averse after large negative shocks and that become more likely to choose the place where industry relations are strong.

The second chapter examines the influence of land supply on housing markets in urban China. The extent to which geographical and man-made land constraints influence housing prices and quantities is explored. Using a sample of 35 cities in China from 2003 to 2012, I find that cities with less naturally available land have experienced greater price appreciation and the quantity response is less in those places. The results imply that geography matters in Chinese housing markets where land is discretely allocated by the government. In cities where there is more land naturally available, the government may be less concerned about the loss of arable land and be more permissive with development. Moreover, my findings imply that the allocation of land use via government decision in China is quasi-exogenous to changes in housing price and quantity, suggesting that decisions by governments about land supply may not be dependent on housing prices.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

The Location of New Manufacturing Firms: How Important Are Agglomeration Economies?

1.1 Introduction

Clustering of firms may be a key driver of job growth and new firm formation (Delgado et al., 2010; Glaeser and Kerr, 2009). As Marshall (1890) points out firms may want to locate near one another because they can benefit from transport cost savings and thick local labor markets. The decline of U.S. manufacturing activity and employment in recent times raises a question whether or not agglomeration economies are still important. Countries like China with cheap labor boost demand for foreign-made intermediate inputs and final goods at the expense of products made in the United States. The introduction of robots and machines reduces the demand for labor. Those trends may weaken the influence of input-output linkages and labor market pooling for the manufacturing sector.

The objective of this paper is to investigate whether agglomerative forces continue to have explanatory power over a time period when the United States experienced a decline in manufacturing activity. I extend previous work on the empirics of agglomeration economies by exploring the determinants of new firm locations in the United States during the period 2004-2011. An advantage of the chosen time period is that it allows the exploration of how the negative macro shocks of the financial crisis of 2008 influence firm location decisions and their response to agglomerative forces.

Clusters of firms arise for many reasons. Natural advantage may account for a portion of geographic concentration. For instance, the location of firms that manufacture petroleum and coal products are likely affected by the location of reserves of fossil fuels. However, geographic concentration is too great to be explained solely by differences in natural resources. Marshall (1890) described three mechanisms of agglomeration. First, the cluster of firms enables them to share large sets of input suppliers and to close their intermediate good custumers. Second, industries using similar types of workers may co-locate so that firms and employees both benefit from locating in a thick labor market. Third, employees may learn knowledge and skills quickly from each other in the industrial cluster. In addition to natural advantages and Marshall's, agglomeration mechanisms institutional factors may affect firms's location decisions. Actions taken by the public sector, in particular, taxes, environmental regulations and incentive programs, are also crucial to the new business (Arauzo-Carod et al., 2010).

Policies that encourage the form of industrial clusters have been largely ignored by policymakers at the federal level in the United States. Economic policies have traditionally focused on either stabilizing the general business environment or supporting individual firms (Porter, 2007). On the one hand, federal economic policy is inclined to monitor macroeconomic stabilization. On the other hand, local government development policy focuses on local benefits. For example, the opening of a new large plant may generate employment growth and productivity benefits in the local area (Greenstone et al., 2010). Policy initiatives aimed at regional level have been given attention in recent decades. The success of firm clusters, like Silicon Valley, has shifted local economic policy to the point where an entrepreneurial cluster has been promoted routinely. Lessons from recent and past crises have emphasized the importance of creating strong urban communities to insulate the local economy from macro shocks. Because of the presence of supplier linkages, labor market pooling and knowledge spillovers, agglomeration effects may help the local economy recover quickly from recession. My study will specifically examine whether the workings of agglomeration economies have become more pronounced after the Great Recession by exploring the determinants of firm births before and after the Great Recession, 2004-2007 and 2008-2011.

My analysis has two main parts. First, I explore the determinants of industry clusters by examining the location of new manufacturing firms in United States over a substantial time period. In particular, I estimate the roles of Marshallian factors and local conditions in generating new firm activity between 2004 and 2011. I use the Reference USA historical business dataset[1], which has only recently become available for researchers to use. I replicate my analysis at the Metropolitan Statistical Area level and at the county level given the concern that industry spillover and local conditions may operate at different geographical units.

Second, I focus on comparing and contrasting new firm creation in the pre- and post-crisis time periods. Previous research discusses on the one hand, how the presence of a strong cluster could make the regional economy more resilient to shocks (Delgado et al., 2015). On the other hand, a cluster could make a region more vulnerable to negative shocks when the shocks propagate among industries (Acemoglu et al., 2013). The financial crisis of 2008 provides an opportunity to investigate whether negative macro shocks would change how firm location decisions respond to the agglomeration effects. A simple approach is explored. Given the richness of the firm-level data set, I am able to divide the analysis into two time periods, 2004-2007 and 2008-2011. The comparison between the

[1] Reference USA website: http://www.referenceusa.com/Home/Home

two time periods allows me analyze if there has been a strengthening or weakening of agglomerative forces during the recent chaotic financial times at the national level.

My main findings can be summarized as follows. Labor market pooling and input-output linkages have the most robust effects, positively influencing agglomeration at all levels of geography. Knowledge spillovers positively affect agglomeration only at the county level. Natural advantages can partially explain the geographic distribution of manufacturing activities. Moreover, I find that corporate taxes discourage firm births but the effects are weaker in more geographically concentrated industries. The comparison of the two periods suggests that there has been a strengthening of agglomerative forces after the Great Recession. One possible explanation is that negative shocks may make new firms more risk averse and that they are more likely to choose the place where the industry relations are strong.

The rest of the paper is organized as follows. In section 2, I discuss the relevant literature. Section 3 presents the location choice model. Section 4 and section 5 describe the data and variables. Section 6 lays out the empirical specification. In section 7, I report and discuss the results. Section 8 concludes.

1.2 Related Literature

A rich empirical literature on agglomeration economies focuses on the determinants of geographical concentration[2]. There are identification issues related with those approaches: the presence of omitted variables and simultaneity. An approach to deal with endogeneity problems was first developed Rosenthal and Strange (2003). They estimate the births of new establishments and their associated employment levels as functions of local industrial characteristics. Results indicate that agglomeration economies attenuate with distance and that industrial organization affects the benefits of agglomeration. Ellison et al. (2010) address identification difficulties by developing two sets of instrumental variables. The results support the empirical relevance of the Marshallian agglomeration factors in that the coefficients on all three mechanisms are positive and significant. Input sharing is the most important agglomeration mechanism. Agglomeration economies have also been found in studies in other countries (Jofre-Monseny et al., 2011; Autant-Bernard, 2006; Guimaraes et al., 2000; Roberto, 2004; Egeln et al., 2004; Wu, 1999). Empirical work shows the evidence that agglomeration effects are stronger in less advanced countries like China, India and Colombia (Chauvin et al., 2013; Combes et al., 2015; Duranton, 2016).

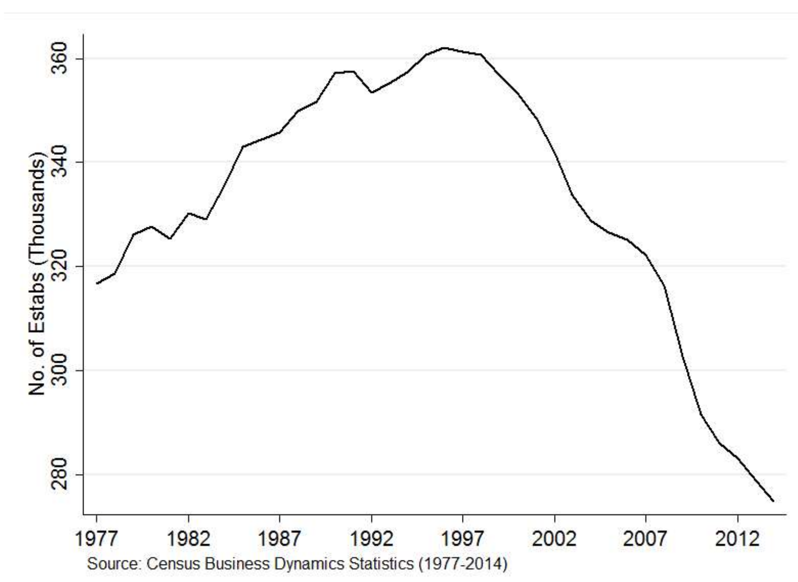
It has long been recognized that natural advantages can also affect the location decisions

[2]See Rosenthal and Strange (2004) and Combes and Gobillon (2015) for review articles.

of firms (Kim, 1999; Ellison and Glaeser, 1999). Ellison et al. (2010) construct an index which reflects agglomeration due to natural advantage based on the 16 natural advantages studied in Ellison and Glaeser (1999). Earlier studies on the effect of taxation have yielded mixed results (Carlton, 1983; Brülhart et al., 2012; Rohlin et al., 2014).

Agglomeration effects may be heterogeneous over time. Many discussions either show how agglomeration effects are becoming less important, as transportation costs have fallen or, instead, how proximity increasingly matters (Duranton, 2016). This paper is closely related to the empirical literature that seeks to determine the relative importance of agglomeration mechanisms. Glaeser and Kerr (2009) study the local determinants of manufacturing firm entry at the city level for the time interval 1976-1999 when the number of manufacturing establishments was increasing, as shown in figure 1. They found evidence that local labor market pooling is strong. Input sharing appears to matter less than labor pooling. However, there has been a steady decrease in manufacturing establishments in the U.S. starting from late 1990s. It would be interesting to understand whether agglomeration effects remain important to the location choice when the manufacturing sector experiences a persistent decline. I extend previous work by exploring the effects of industrial externalities, taxes and natural advantage on new firm location decisions during the period of 2004-2011. To complement the existing urban literature, my paper aims to investigate the importance of agglomeration effects before and after the 2008 financial crisis, in particular to examine whether the nation-wide negative shock has any impact on the workings of agglomeration mechanisms at the local level.

Figure 1: Number of Manufacturing Establishments in the US (1977-2014)



1.3 A Model of Location Choice by New Firms

In this section, I explain a simple model in which geographical concentration is the result of random profit-maximizing location decisions made by new firms. Industry-specific spillovers and natural advantages lead firms to cluster together.

Firm i chooses from J options correspond to the area that will yield the highest expected, the profit derived by firm i if it locates at area j is given by (Carlton, 1983; McFadden, 1973; Arauzo-Carod et al., 2010; Bhat et al., 2014):

$$\pi_{ij} = \gamma z^{'}_{ij} + \varepsilon_{ij}, \quad i = 1, ..., N; j = 1, ..., J, \tag{1}$$

where $z^{'}_{ij}$ represents a vector of explanatory variables and ε_{ij} is an error term that is iid extreme value. The probability that the firm n chooses alternative is:

$$p_{ij} = \frac{exp(\gamma z^{'}_{ij})}{\sum_{j=1}^{J} exp(\gamma z^{'}_{ij})}. \tag{2}$$

Given data on firms' choices, γ can be estimated by maximizing the log likelihood function (Guimaraes et al., 2003):

$$\log\ L_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{J} q_{ij}\ \log\ p_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{J} q_{ij}\ \log\ \frac{exp(\gamma z^{'}_{ij})}{\sum_{j=1}^{J} exp(\gamma z^{'}_{ij})}, \tag{3}$$

where $q_{ij} = 1$ in case firm i choose location j and $q_{ij} = 0$ otherwise.

Estimation of γ is complicated in the presence of agglomeration effects. To see this, suppose firm i and firm k affect each other's location decisions simultaneously. These effects are difficult to identify in the firm level regression above. δ represents the effect of firm k on firm i's location decisions when both firms choose to locate in geographic unit j :

$$\log\ L_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{J} q_{ij}\ \log\ p_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{J} q_{ij}\ \log\ \frac{exp(\gamma z^{'}_{ij} + \delta s_{ik})}{\sum_{j=1}^{J} exp(\gamma z^{'}_{ij} + \delta s_{ik})}, \tag{4}$$

suppose s_{ik} is omitted from the regression, and the relation between $x^{'}_{i}$and s_{ik} is given by $s_{ik} = \theta z^{'}_{ij}$. Equation (4) can be written as:

$$\log\ L_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{J} q_{ij}\ \log\ p_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{J} q_{ij}\ \log\ \frac{exp(\gamma z^{'}_{ij} + \delta\theta z^{'}_{ij})}{\sum_{j=1}^{J} exp(\gamma z^{'}_{ij} + \delta\theta z^{'}_{ij})}. \tag{5}$$

Endogeneity bias introduced by agglomeration is difficult to address because most quasi-experimental sources of variation will impact both firm i and k. Moreover, the direction of the bias is not clear because the sign of γ can be positive or negative. The sign of δ would be positive in the case that firms can benefit from each others when they choose to locate in the same areas. In contrast, firms may choose to avoid locating close to their competitors if they suffer a decline in market share. The sign of δ is likely to be negative when the cost of clustering overweighs the benefits. One way of dealing with the issue is by moving toward aggregate territorial units. Most recent research on location choices has been based on count data models. A count model considers territorial location as the unit of analysis and can be derived as an aggregatelevel reduced form. Second, firm-level estimation generally uses very few firm characteristics because of the unavailability of such data (Arauzo-Carod et al., 2010). These issues can turn aggregate territorial-level regression into the preferred specification. Guimaraes et al. (2003) assume that individual decisions are based on a vector of choice specific variables common to all firms, $z_{ij} = z_j$:

$$\log\, L_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{J} q_{ij}\, \log\, p_{ij} = \sum_{j=1}^{J} n_j\, \log\, p_j = \sum_{j=1}^{J} n_j\, \log\, \frac{exp(\gamma z_j^{'} + \bar{\delta}\theta z_j^{'})}{\sum_{j=1}^{J} exp(\gamma z_j^{'} + \bar{\delta}\theta z_j^{'})} \tag{6}$$

$\bar{\delta}\theta z_j^{'}$ can be replaced with regional fixed effects. Guimaraes et al. (2003) proved that log likelihood coefficients can be equivalently estimated using the Poisson regression with exponential mean function

$$E(n_j) = exp(\gamma z_j^{'}), \tag{7}$$

where $E(n_j)$ is the count of new births in industry i that locate in geographical j. Poisson models are particularly useful when a highly disaggregated territorial level is used (Arauzo-Carod, 2008). Because the area of each unit is small, a large number of these areas is likely to not receive any new establishments. Poisson models are ideally structured to deal with the zero problems.

1.4 New Manufacturing Firms

In recent years, the increasing availability of firm-level data has enabled scholars to access data at very detailed geographical units. The manufacturing sample that I use is retrieved from the ReferenceUSA Historical Business Database. This firm-level database contains the industry of each firm[3] and its location, employment size, corporate structure and more, tracing the firm information from its beginning year. In my empirical work, I define the dependent variable as the count of firms in the manufacturing sector established between 2004 and 2011 by industry and location. The industry definition that I use corresponds to the three or four digit level of the 2002 North American Industry Classification system. I begin with 2600 MSA-industry pairs that are formed by using the top 50 Metropolitan Statistical Areas in the United States which have population above 1.1 million in 2010. Alternatively, in order to investigate agglomeration sources that are across small geographical units within dense areas, I construct a smaller sample consisting of 299 counties which located in the top 35 MSAs with population above 1.8 million in 2010.

Table 1 includes the five MSAs, counties and industries with the highest number of new manufacturing firms over the time period 2004-2011. Table 2 documents distributions of manufacturing firm type. 92.47% of firms are single locations. This paper only focuses on single locations. I report the mean annual entry counts and entry employments of new firms over the 2004-2011 in table 3. Figure 2 presents the distribution of establishment entry sizes. Over three-fourths of new firms begin with four or fewer employees.

Figure 3 presents the distributions of the dependent variables at the MSA-industry level and county-industry level. Firm entry distributions are highly skewed since many MSA-industry and county-industry observations experience very limited entry. OLS regression would be inappropriate to use in the estimation where the data-generating process is so skewed. Previous empirical work has dealt with the excessive number of zeros, in one of two models: the Tobit model and the Count model. The Tobit model is designed to estimate the relationships between variables when the dependent variable is either left-censored or right-censored. In some data sets, we cannot observe values above or below some threshold because of a censoring or truncation mechanism. Tobit models allow for these cases. However, Tobit models have the limitation that they consider the zero outcome to be the result of censoring, whereas a zero outcome is a natural outcome variable in the firm-level location data (Rocha, 2008). Count data models, including Poisson and Negative Binomial models, consider territory as the unit of analysis. Ideally, small geographical units are preferred because large geographic units contain heterogeneity within themselves (Guimaraes et al., 2003). The count data approach allows for large sets of location choices with frequent zero outcomes. The problem with Poisson regression models is that count data frequently suffers from overdisperion (variance greater than the mean) which violates the Poisson assumption of equal mean and variance. A common practice is to adopt the Negative Binomial model which does not impose the restriction of equal mean and vari-

[3]RefUSA is a data set of establishments, my paper study new firms that are single-location only.

Table 1: New Manufacturing Firms in Top 50 MSAs (2004-2011)

MSAs	New firm count
Panel A. five MSAs with the highest number of new firms	
Los Angeles-Long Beach-Santa Ana, CA	18,676
New York-Newark-Edison, NY-NJ-PA	17,317
Chicago-Naperville-Joliet, IL-IN-WI	9,332
Dallas-Fort Worth-Arlington, TX	7,056
Miami-Fort Lauderdale-Miami Beach, FL	6,842

Counties	New firm count
Panel B. four counties with highest number of new firms	
Los Angeles, CA	13,835
Cook, IL	4,865
Orange, CA	4,841
Harris, TX	4,499
Dallas, TX	3,827

Industry	New firm count
Panel C. four industry with the highest number of new firms	
Other miscellaneous manufacturing (3399)	27,195
Printing and Related Support Activities(3231)	23,301
Household and Institutional Furniture and Kitchen Cabinet Manufacturing (3371)	14,504
Bakeries and Tortilla Manufacturing (3118)	14,110
Machine Shops; Turned Product; and Screw, Nut, and Bolt Manufacturing (3327)	7,788

Sources: Reference USA Business Historical Database (2004-2011)

Table 2: Distribution of Manufacturing Firm Types (Entire Country)

Total Firms	Single Location	Percent	Branch	Percent
278,601	257,610	92.5	20,991	7.5

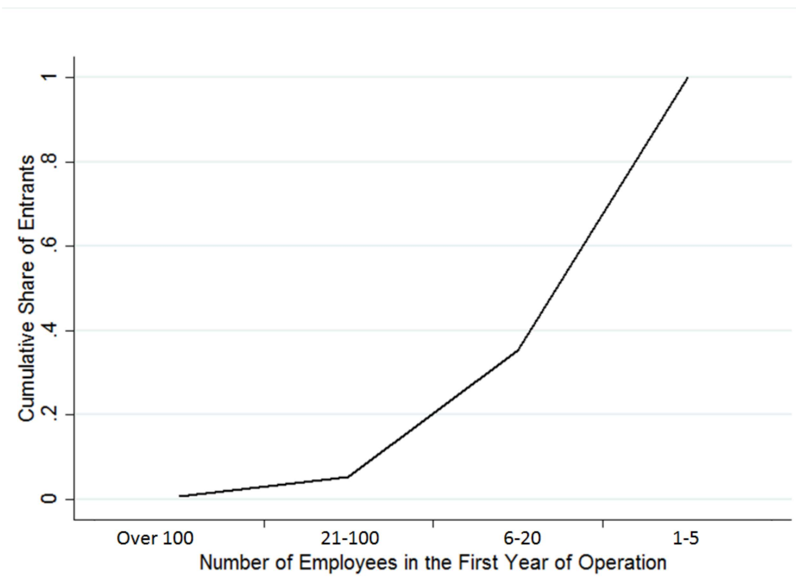
Sources: Reference USA Business Historical Database (2004-2011)
Notes: Only single locations are considered in this paper.

Table 3: Manufacturing Firm Entry (Entire Country)

Mean Annual Entry Counts	32,348
Counts by Firm Size	
1-4 Employees	64.2%
5-19 Employees	29.8%
20-99 Employees	4.5%
101+ Employees	1.5%

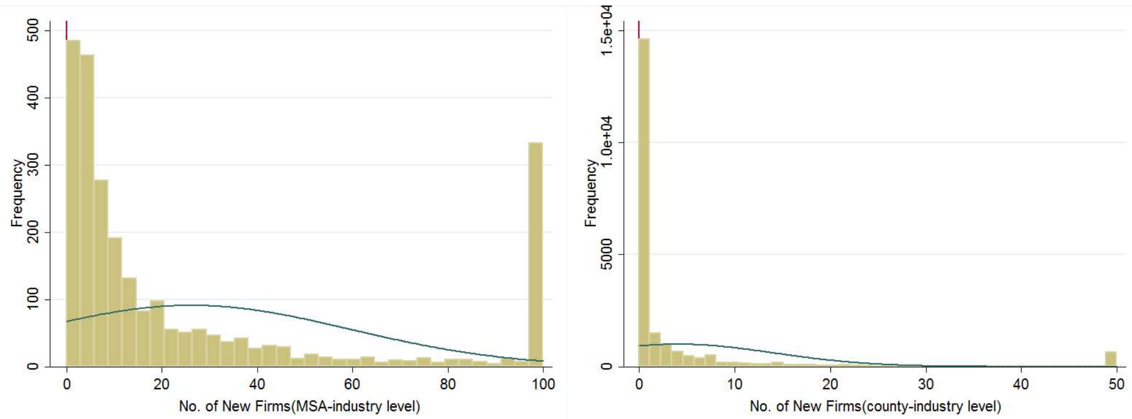
Sources: Reference USA Business Historical Database (2004-2011)

Figure 2: Distribution of Firm Entry Size (2004-2011)



Sources: Reference USA Business Historical Database (2004-2011)

Figure 3: Distribution of Dependent Variables



(a) MSA-industry level

(b) County-industry level

ance, and so the Negative Binomial is my preferred estimation technique. For comparison, I also report the results for Tobit models in appendix A tables 12-13.

1.5 The Determinants of Industrial Location

The goal of this section is to describe how I measure of the determinants of firm location. My strategy is to use the Negative Binomial model to regress counts of new births on proxies for Marshallian factors: input sharing, labor market pooling, and knowledge spillovers. I also provide controls for natural advantages and local government policies. Summary statistics are provided in Table 4.

1.5.1 Agglomeration Theories

Agglomeration economies are probably the most studied determinants of industrial location and their measures can be elusive. Marshall (1890) discuss three main sources of agglomeration economies: labour market interactions, linkages between intermediate- and final-goods suppliers, and knowledge spill-overs. Duranton and Puga (2004) distinguish three types of micro-fondation of urban agglomeration economies based on sharing, matching, and learning mechanisms[4]. My primary goal is to assess the importance of Marshall's theories of agglomeration to the manufacturing sector in the United States. In the urban

[4]Sharing is about the possible gains from the wider variety of input providers, the diversity of local goods, the division of labour, or risks. Matching is about the probability of finding another party such as a worker, an employer, a supplier and the greater quality of the match with that party. Learning is about the generation, diffusion, and accumulation of knowledge (Duranton, 2015).

Table 4: Summary Statistics

	Mean	S.D.	Min	Max
No. of firms (MSA-industry level)$_{04-11}$	58.35	193.00	0	3,244
No. of firms (MSA-industry level)$_{04-07}$	31.87	108.67	0	1,925
No. of firms (MSA-industry level)$_{08-11}$	26.48	85.80	0	1,411
No. of firms (county-industry level)$_{04-11}$	8.77	49.01	0	2,119
No. of firms (county-industry level)$_{04-07}$	9.27	38.62	0	1,240
No. of firms (county-industry level)$_{08-11}$	7.70	29.27	0	902
No. of employees (MSA-industry level)$_{2002}$	2,668	5,840.01	0	80,838
No. of employees (MSA-industry level)$_{2007}$	1,803	4614.123	0	57,567
No. of employees (county-industry level)$_{2002}$	340	1,741.01	0	71,623
No. of employees (county-industry level)$_{2007}$	265	1,400.54	0	52,636
Coporate tax rate(%)$_{2002}$	5.62	2.68	0	9.99
Coporate tax rate(%)$_{2007}$	6.33	2.09	0	9.99
Electricity price (cents per kilowatthour)$_{2002}$	5.00	1.51	3.04	9.37
Electricity price (cents per kilowatthour)$_{2007}$	6.84	1.97	3.95	13.03
Coal mining production (000 tons)$_{2002}$	71.99	599.12	0	7,027
Coal mining production (000 tons)$_{2007}$	74.29	519.95	0	4,488

Notes: No. of firms refers to number of new manufacturing firms, data is retrieved from Reference USA Business Historical Database (2004-2011). No. of empolyees refers to number of existing employees in year 2002, data are drawn from the Economic Census: *Manufacturing: Industry Statistics for the States, Metropolitan and Micropolitan Statistical Areas, Counties, and Places*. Mean and standard deviations for No. of firms and No. of employees are measured across industry and regional level.

economics literature, the strongest evidence by far is for labor market pooling. The evidence on input sharing is mixed. The presence of intermediate good customers is likely to encourage new firm births, while the presence of input sources is likely to encourage the birth of new plants by old firms (Rosenthal and Strange, 2004). Knowledge spillovers have been tricky to measure and may have somewhat weaker effects. Intellectual sharing may be better captured by occupation correlations (Porter, 1990) than by patent citations. In the following subsections, I briefly discuss the Marshallian mechanisms and the metrics I construct to capture industrial spillovers.

Input shares: Some of mechanisms of agglomeration that Marshall discusses include input sharing–firms locate near one another to share a large base of suppliers or to be closer to intermediate good customers. A concentration of firms enables them to reduce the cost of obtaining inputs and shipping goods to customers. Because of technologies and quality of goods, there has been a remarkable decline in transportation costs in the past decades (Glaeser and Kohlhase, 2004). One of objectives of this paper is to access whether supplier-consumer relationships remain important when transportation costs are likely decreasing.

To test the importance of the mechanism, I use 2002 and 2007 Input-Output Accounts published by the Bureau of Economic Analysis (BEA) to measure the extent that industries buy and sell from one another. The input-output tables provide information on the commodity inputs that are used by industries and commodities produced by industries. I construct two sets of weight following previous work (Jofre-Monseny et al., 2011):

$$S_{ij}^{I} = \frac{inputs_{i \leftarrow j}}{total\, inputs_i}, \tag{8}$$

$$S_{ij}^{O} = \frac{outputs_{i \rightarrow j}}{total\, outputs_i}, \tag{9}$$

where S_{ij}^{I} is defined as the share of industry i's input that come from industry j (including those in the agriculture and the services sectors), S_{ij}^{O} is defined as the share of industry i's output that is sold to industry j. The shares range from zero to one.

Based on the weights described in (9) the industry that most intensely relies on input suppliers is motor vehicle manufacturing (NAICS 3361) which obtains 59.1$\%$ of its inputs from producers of motor vehicle bodies, trailers and parts (NAICS 3362). The second highest input share value of S_{ij}^{I} is 0.485, which represents 48.5$\%$ of inputs that come to the manufacturing of pulp, paper and paper board mills (NAICS 3221) comes from the manufacturing of converted paper products (NAICS 3222). The highest value of the output shares S_{ij}^{O} is 0.503 for manufacturing of motor vehicle bodies, trailers and parts (NAICS 3362), which represents 50.3$\%$ of their output is sold to the motor vehicle manufacturing (NAICS 3361). The second highest value is 0.422, which show the producers of resin, rubber and artificial fibers (NAICS 3252) sell 42.2$\%$ of their outputs to plastics and rubber products manufacturing (NAICS 3260). Based on these two sets of shares I construct the variables $input_{ig}$ and $output_{ig}$:

$$input_{ig} = \sum_{j \neq i} (S_{ij}^{I} \cdot E_{gj}), \tag{10}$$

$$output_{ig} = \sum_{j \neq i} (S_{ij}^{O} \cdot E_{gj}). \tag{11}$$

The bracketed term in equation (10) multiplies the national share of industry i's input that come from industry j (S_{ij}^{I}) with industry j's employment in the location g (E_{gj}). Industries that have stronger supplier relationships with industry i are given higher weights. Employment data are drawn from the Economic Census[5]. Data sets have been published every five

[5]Manufacturing: Industry Statistics for the States, Metropolitan and Micropolitan Statistical Areas, Counties, and Places

years (2002, 2007, 2012, etc.). I report the descriptive statistics of employment in table 4. By summing across industries, I obtain $input_{ig}$ which measure the local employment in the industries that provide inputs to industry i's. I apply the same methodology to construct the variable $output_{ig}$ where industries have stronger intermediate good customer relationships are given higher weights. The construction of $output_{ig}$ measures the local employment in the industries that are industry i's buyers.

Labor market pooling: The location of manufacturing firms might become less dense because of low transport costs for goods. However, moving people is still expensive (Glaeser and Kohlhase, 2004). Labor may be the most important factor for any new firm. Many industries require specialized workers. The location of new firms could be a function of the concentration of other firms because there are gains from a thick labor market. Marshall argued about the risk-sharing properties of a thick labor market. Workers can be better shielded from firm-specific negative shocks by moving across firms and industries (Diamond and Simon, 1990; Overman and Puga, 2002). Meanwhile, firms can experience more efficient matches and be more productive when accessing larger labor pools. These properties suggest that firms that use similar workers may have advantages if they locate near one another.

The occupational similarity index is intended to capture the importance of labor pooling. The National Industrial-Occupation Employment Matrix 2002 and 2007 (NIOEM) is the source for occupation data. The NIOEM published by the Bureau of Labor Statistics catalogues occupational employment patterns across industries with 462 occupations. Following previous work (Duncan and Duncan, 1955; Jofre-Monseny et al., 2011), the variable $occupationalsimilarity_{ij}$ measures the extent to which industry i and j use similar types of labor:

$$occupational\ similarity_{ij} = 2/\sum_{o} |\frac{L_{ki}}{L_i} - \frac{L_{kj}}{L_j}|, \tag{12}$$

where $\frac{L_{ki}}{L_i}$ denotes the share of occupation k in the industry i. The more similar are workers that the two industries use, the smaller the absolute differences between the share of occupation k in the industry i and the share of occupation k in the industry j, and the larger the value of $occupational\ similarity_{ij}$.

To increase the weights assigned to the most similar industries, I sort in descending order all industries based on the occupational similarity with industry i and only consider the six closest industries (these are all within one standard deviation above the mean), Following previous work (Jofre-Monseny et al., 2011) I define:

$$S_{ij}^{os} = 0 \quad if\ rank > 6th, \tag{13}$$

$$S_{ij}^{os} = \frac{occupational\ similarity_{ij}}{\sum_{rank=1}^{6} occupational\ similarity_{ij}} \quad if\ rank \leq 6th. \tag{14}$$

Industrial machinery manufacturing (NAICS 3332) and other general purpose machinery manufacturing (NAICS 3339) have the most similar employment pattern among industries pairs. Based on the weights the variable $labor_{ig}$ is constructed as:

$$labor_{ig} = \sum_{j \neq i} (S_{ij}^{os} \cdot E_{gj}), \tag{15}$$

where $labor_{ig}$ measures local employment in the industries that use similar type of workers with industry i.

Knowledge spillovers: Knowledge spillovers could be a function of clustering because there are gains from people being able to interact. Marshall considered that employees learn skills and knowledge easily from each other in an industrial cluster. However, knowledge spillovers are difficult to identify. In the literature, the most direct test of knowledge spillovers is provided by patent citations showing that firms at knowledge-intensive industries are more likely to cite other firms who are spatially closer (Jaffe et al., 1993; Agrawal et al., 2008, 2010), although the implied effect tends to be weak (Glaeser and Kerr, 2009; Ellison et al., 2010).

Research on knowledge spillovers in my paper has been limited given imperfect measures of intellectual spillovers and unavailability of national patent data classified by industry. The source of data on knowledge spillovers I use is based on Ellison et al. (2010) patent matrix which captures industry i citations to technologies associated with industry j, and vice versa. They constructed measures of intellectual spillovers across an industry pair using the NBER Patent Database[6].

The constructed patent matrix from Ellison et al. (2010) corresponds to the 1987 Standard Industrial Classification (SIC). I use the concordance between 1987 SIC and the 2002 NAICS provided by the Census Bureau to convert the 1987 SIC patent matrix to the 2002 NAICS matrix. $patentin_{i \leftarrow j}$ represents industry i cite technologies from industry j and $patentout_{i \rightarrow j}$ represents industry i's technologies are cited by industry j. In a manner

[6]The NBER Patent Data file contains records for all patents granted by the United States Patent and Trademark office (USPTO) from January 1975 to December 1999. The USPTO classifies patents data by technology categories rather than by industries.Ellison et al. (2010) develop concordances between the USPTO classification and SIC3 industries

analogous to the weights I defined for my measures of labor market pooling, I only consider the four closest values (which fall within one standard deviation above the mean) with industry i. If rank $>$ 4th:

$$S_{ij}^{pi} = 0, \quad S_{ij}^{po} = 0, \tag{16}$$

and if rank $\leq$ 4th:

$$S_{ij}^{pi} = \frac{patentin_{i \leftarrow j}}{\sum_{rank=1}^{4} patentin}, \quad S_{ij}^{po} = \frac{patentin_{i \rightarrow j}}{\sum_{rank=1}^{4} patentout} \tag{17}$$

Based on the set of weights I construct the variables $citing_{ig}$ and $cited_{ig}$ which are measures of local employment that share knowledge with industry i:

$$citing_{ig} = \sum_{j \neq i} (S_{ij}^{I} \cdot E_{gj}), \tag{18}$$

$$cited_{ig} = \sum_{j \neq i} (S_{ij}^{O} \cdot E_{gj}), \tag{19}$$

where industries that cite more patents in their production processes are given higher weights. Hence, $citing_{ig}$ and $cited_{ig}$ are measures of the local employment in the industries that share knowledge and ideas with industry i.

1.5.2 Natural advantage

In addition to Marshallian spillovers, empirical work on firm clustering often looks at a simpler alternative: an industry may be concentrated if firms choose the locations that have natural advantages (Kim, 1999; Ellison and Glaeser, 1999). Previous work finds that only one-fourth of the propensity to cluster can be attributed to natural advantage (Ellison and Glaeser, 1999). A simple way to identify effects of natural advantage on firm clustering is to regress the number of firms in a given industry at the county-level on the county's resource endowmen (Kim, 1999). However, this approach does not consider whether or not an industry is sensitive to the cost of a particular input (Ellison and Glaeser, 1999). For instance, coal products manufacturing is more sensitive than pharmaceutical manufacturing to the location of coal mining sites. To better measure natural advantages, I multiply state (county)-level input variables (e.g. coal mining production) by the industry ratio which reflects the intensity of input use (the share of industry i's input that comes from coal mine industry). Two variables are designed to reflect the costs of two common inputs for manufacturing: $coal\,mining\,production \times coal\,use\,ratio$ and $electricity\,price \times electricity\,use\,ratio$. I obtain the data for resource endowments from U.S. Energy Information Administration (EIA)[7]. Data for input use ratio is retrieved from the US National Input-Output Accounts.

1.5.3 Tax impacts

The effect of taxation on industrial location is an issue that has been investigated by scholars. According to earlier studies, taxation exerts a negative effect on the location of firms (Brülhart et al., 2012; Jofre-Monseny and Solé-Ollé, 2012; Rohlin et al., 2014). I focus on state corporate taxes. My data source for taxes comes from the Tax Foundation[8], which provides state corporate tax rates and brackets.

Brülhart et al. (2012) presents evidence that agglomeration forces can offset differences in corporate taxes as determinants of firm location. The authors use an interaction term between local corporate tax rates and a measure of agglomeration to estimate the sensitivity of firm location to local taxes: $tax \times EGindex$. The Ellison-Glaeser (EG) index is a measure of agglomeration which identifies the concentration of industry. Ellison and Glaeser (1997) define $EG - index$:

$$\gamma_j^{EG} = \frac{\sum_{i=1}^{M}(s_i - x_i)^2 - (1 - \sum_{i=1}^{M} x_i^2)H_j}{(1 - \sum_{i=1}^{M} x_i^2)(1 - H_j)}, \tag{20}$$

where s_i is the share of industry j's employment in area i, x_i is the share of total employment in area i, $Herfindahl\, index\; H_j = \sum_{k=1}^{N} z_k^2$, z_k is the size of the establishment k of industry j. A positive estimated coefficient on the interaction term, $tax \times EGindex$, implies that location decisions of firms in more clustered industries are less sensitive to tax differences. A negative coefficient implies that firms in industries with high $EG - indexes$ are more sensitive to taxes. I assemble $EG - index$ from a variety of sources: Manufacturing: Industry Statistics for the States, Metropolitan and Micropolitan Statistical Areas, Counties, and Places (s_i); Concentration Ratio ($Herfindahlindex$); County Business Pattern Data (x_i). When comparing the values I compute for the $EG - index$, I find that Computer and peripheral equipment manufacturing (NAICS 3441) is relatively dispersed, with the lowest $EG - index$ (EG=-0.21). Conversely, metalworking machinery manufacturing industry (NAICS 3335) is the industry with a highest degree of geographical concentration (EG=0.032).

[7]EIA website: https://www.eia.gov/

[8]Tax Foundation website: http://taxfoundation.org/

1.6 Empirical Specification and Identification Issues

I now present my empirical specification of how industry spillovers may contribute to firm births at different geographical scales for the time interval 2004-2011. Negative Binomial regressions have been performed using firm level data aggregated to the MSA and county level. I begin by characterizing MSA-level traits with only Marshallian factors being considered as explanatory variables:

$$N_{i,g} = \beta_1 \cdot Marshallian_{i,g} + \beta_2 \cdot E_{i,g} + \alpha_i + \alpha_g + \varepsilon_{i,g}, \tag{21}$$

where the dependent variable N_{ig}, is the count of new firm creations in industry i and geographical unit g between 2004 and 2011. Marshallian factors include: (1) input-output linkages $input_{ig}$, $output_{ig}$; (2) labor market pooling $labor_{ig}$; (3) knowledge spillovers $citing_{ig}$, $cited_{ig}$. I further control for the pre-existing number of own establishments in each MSA, $E_{i,g}$ and two additional controls, industry fixed effects (α_i) and MSA fixed effects (α_g). The term $\varepsilon_{i,g}$ corresponds to the error term.

As mentioned earlier, a potential concern with the specifications above is likely omitted variables. The estimation would be biased if omitted variables are correlated with variables representing geographical characteristics, which could lead to reverse causality. Marshallian spillovers may be the result and not the cause of industry clustering. To address the issue of simultaneity bias, I estimate the count of new firms by industry and location between 2004 and 2011 as a function of pre-determined variables. Therefore, the explanatory variables correspond to 2002 to avoid potential simultaneity. Some omitted natural advantage variables are still likely correlated with Marshallian factors. The inclusion of location-specific fixed effects partially addresses this issue. The term α_i corresponds to industry fixed effect and α_g corresponds to location fixed effect. Given the aforehand issues, I interpret estimates as partial correlations rather than as causal effects throughout the paper.

I now turn to the county-level analysis. Agglomeration factors may perform differently at different geographic scales. The application of count model for highly aggregated regions poses a problem in that large geographic units may contain heterogeneity within themselves. In practice, small geographic units are preferred because some factors are thought to take place at the local level (Guimaraes et al., 2003). Information on firm characteristics of small territorial units is not usually available with such a degree of detail. In this respect, the existence of a richer dataset, the RefUSA historical business data allows the estimation of location choices aggregated to the county level as well as to the MSA-level.

Data for natural endowments and tax rates are available at the county or state level. Hence, county-level analysis allows me to include additional variables which reflect firm location choices due to natural advantage, taxation and its interaction term with the EG-index. With the control of county-industry employment, county employment, fixed effects, the estimation can be written as:

$$
\begin{aligned}
N_{i,g} = \beta_1 \cdot Marshallian_{i,g} + \beta_2 \cdot T_g + \beta_3 \cdot T_g \cdot EG_i + \beta_4 \cdot Natural\,Advantage_{i,g} \\
+ \beta_5 \cdot \log\, emp_{i,g} + \beta_6 \cdot \log\, emp_g + \alpha_i + \alpha_g + \varepsilon_{i,g}.
\end{aligned}
\tag{22}
$$

where the dependent variable N_{ig}, is the count of new firm creations by industry i and by county g between 2004 and 2011. Marshallian factors include: (1) input-output linkages $input_{ig}$, $output_{ig}$; (2) labor market pooling $labor_{ig}$; (3) knowledge spillovers $citing_{ig}$, $cited_{ig}$. Local corporate ax rate T_g is expected to be negatively correalted with firm births. The interaction term between tax rates and a measure of agglomeration is designed to estimate the sensitivity of firm location to local taxes: $tax \times EGindex$. Natural advantage include two variables to reflect the costs of common inputs for manufacturing: $coal\,mining\,production \times coal\,use\,ratio$ and $electricity\,price \times electricity\,use\,ratio$. I additionally include the overall employment level $\log\, emp_g$ in order to control for the so-called urbanization economies and further control for the pre-existing number of own employments in each county, $E_{i,g}$ and two additional controls, industry fixed effects (α_i) and county fixed effects (α_g). The term $\varepsilon_{i,g}$ corresponds to the error term.

Another concern with the specification is the issue of scale effect that a bigger city is expected to have more births than a smaller area. Bartik (1985) emphasized geographical units with more available land are more likely to be chosen. Therefore I performed a robustness test, where I consider an alternate specification for the dependent variable, firm births are normalized by the land area of the spatial unit. The results can be found in Appendix B Tables 14-15.

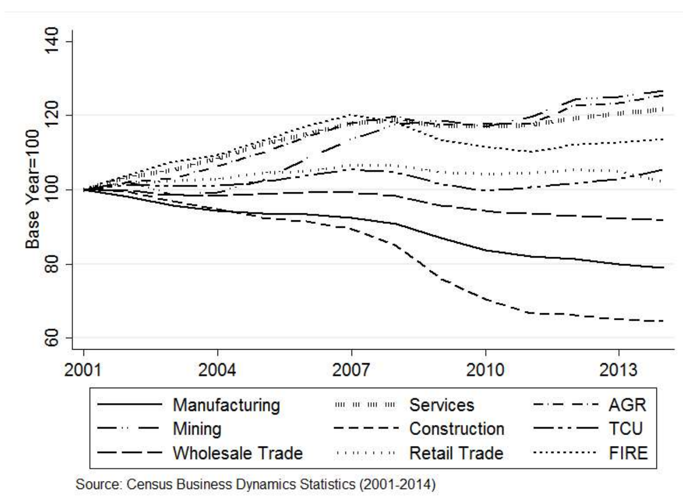
Instead of analyzing all years simultaneously, it is possible to break them down into two periods. The last specification to be highlighted is the regressions presented in equations (22) and (23). I estimate the two time periods separately in order to compare and contrast the effect of agglomeration before and after the Great Recession. The Great Recession had great impacts on the manufacturing sector. It was the third most impacted sector followed by the construction sector and FIRE (finance, insurance, and real estate), as shown in Figure 4. It is worth noting that there is a regional heterogeneity in the decline of new firms after the Great Recession. Table 5 shows from 2004-2007 to 2008-2011, the number of new manufacturing firms decreased by 19.73% in the Midwest region, followed by 16.95% in the Northeast region, the Western region experienced the lowest decline between the two periods. A sharp drop in new manufacturing births does not necessarily mean agglomeration forces lose their function in the local economy. Agglomeration economies could mitigate some of the effects of recessions so that firms choose to locate near clustering areas. An alternative hypothesis is that the workings of agglomerative effects become weaker during the recession. Cluster specialization may propagate negative shocks among related industries, so that firms may choose to avoid areas where industrial clustering is strong. It would be interesting to look at whether negative macro shocks would change the firm location decisions respond to the agglomeration effects. The financial crisis of 2008 provides an opportunity to investigate the workings of agglomeration effects on the local economy. A few steps are taken to generate balanced data sets for each period 2004-2007 and 2008-2012. The comparison between the two data sets allows the discussion of the potential effects of agglomeration externalities over time.

Table 5: Regional Manufacturing Firm Entry (Entire Country)

	2004-2007	2008-2011	$\Delta\%$
Midwest	29,959	24,048	-19.73%
Northeast	24,288	20,171	-16.95%
South	48,555	41,367	-14.8%
West	35,736	33,039	-7.54%

Sources: Reference USA Business Historical Database (2004-2011)

Figure 4: Number of Establishments, by Sectors (2001-2014)



AGR: Agriculture, Forestry, and Fishing
TCU: Transportation, Communication, and Public Utilities
FIRE: Finance, Insurance, and Real Estate

1.7 Results

As I discussed in the introduction, the main aims of this paper are to examine the determinants of firm locations and to compare the different intensities at different geographical units.

I first report and discuss the results at the MSA level. Note that the Marshallian forces are measured in logs, so that the coefficient estimates can be interpreted as elasticities given the Poisson exponential mean specification. The regression results for are shown in table 6. The estimates reported in the first column of table 6 imply that a 1% increase in MSA employees in industries that provide the inputs to industry i increase new firms in industry i by 0.057%. Likewise, that a 1% increase in MSA employees in industries that have intermediate good customer relation to industry i increase new firms creation in industry i by 0.125%. I find the statistically significant evidence of the existence of intermediate good customer relationship, but the presence of input suppliers is weaker. Column 2 finds that labor market pooling is the strongest explanatory variable among the Marshallian factors. Increasing by 1% the employees in industries that use similar workers as those used by industry i is associated with a 0.128% increase in new firms. Labor is an important factor for any new firm, particularly important at highly aggregated geographical units. The results in column 3 indicate that knowledge spillovers have weak correlations with firm location choices. The results imply that knowledge spillovers do not seem to be a driver of clustering in the 2004-2011 time period. However, my findings do not mean these effects are not ever important. Porter (2007) emphasizes knowledge and idea sharing between workers may be better captured by measuring occupation relations. Column 4 reports the regression results obtained when all Marshallian factors are considered simultaneously. The results in table 6 provide suggestive evidences for the importance of labor market pooling and input sharing, but the evidence here is weaker for the knowledge spillovers. Table 7 provides the estimated marginal effects evaluated at the means of the independent variables.

Table 6: Estimations of Mfg. Entry Counts. Negative Binomial Regression, MSA-Industry Level

Dependent Variable	No. of New Firms (2004-2011)			
	(1)	(2)	(3)	(4)
ln Input_{2002}	0.057			0.006
	(0.043)			(0.056)
ln Output_{2002}	0.125***			0.095***
	(0.029)			(0.025)
ln Labor_{2002}		0.128***		0.091**
		(0.044)		(0.046)
ln Citing_{2002}			0.037	0.007
			(0.028)	(0.026)
ln Cited_{2002}			0.019	0.001
			(0.030)	(0.032)
Control				
No. of Establishemnts in $\text{MSA-industry}_{2002}$	Y	Y	Y	Y
Industry fixed effects	Y	Y	Y	Y
MSA fixed effects	Y	Y	Y	Y
No. of Industries	52	52	52	52
No. of MSA	50	50	50	50
Observations	2600	2600	2600	2600

Notes: Estimated model: $N_{i,g} = \beta_1 \cdot Marshallian_{i,g} + \beta_2 \cdot E_{i,g} + \alpha_i + \alpha_g + \varepsilon_{i,g}$. Includes top 50 MSAs (based on 2000 census population). Robust standard errors (in parentheses) are clustered at MSA level. $*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

Table 7: Estimations of Mfg. Entry Counts. Marginal Effects at the Means (Negative Binomial Regression), MSA-Industry Level

Dependent Variable	No. of New Firms	(2004-2011)		
	(1)	(2)	(3)	(4)
ln Input_{2002}	0.767			0.082
	(0.586)			(0.757)
ln Output_{2002}	1.689***			1.275***
	(0.393)			(0.332)
ln Labor_{2002}		1.723***		1.229**
		(0.592)		(0.623)
ln Citing_{2002}			0.503	0.092
			(0.381)	(0.351)
ln Cited_{2002}			0.250	0.001
			(0.401)	(0.424)
Control				
No. of Establishemnts in MSA-industry_{2002}	Y	Y	Y	Y
Industry fixed effects	Y	Y	Y	Y
MSA fixed effects	Y	Y	Y	Y
No. of Industries	52	52	52	52
No. of MSA	50	50	50	50
Observations	2600	2600	2600	2600

Notes: Estimated model: $N_{i,g} = \beta_1 \cdot Marshallian_{i,g} + \beta_2 \cdot E_{i,g} + \alpha_i + \alpha_g + \varepsilon_{i,g}$. Includes top 50 MSAs (based on 2000 census population). Robust standard errors (in parentheses) are clustered at MSA level.$*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

Industries' spillovers may perform at different intensities at different geographical units. Table 8 and 9 report the county-level estimations. Table 9 reports estimated marginal effects associated with the negative binominal regressions. Column 1 of table 8 and 9 include three Marshall agglomeration measures as well as natural advantage and tax effects. MSA fixed effects control for a wide range of metropolitan characteristics that might affect firm births. One can assume that firms first choose which MSA to locate and then decide in which county to locate within the chosen MSA. Such a structure produces random components correlated between counties within a given MSA (Jofre-Monseny et al., 2011; Combes and Gobillon, 2015). The presence of input suppliers and industrial customers seems to drive location decisions of manufacturing startups. Labor market pooling appears insignificant in this specification. It is not hard to imagine that workers are reluctant to live in one MSA and work in another. But workers may be willing to live and work across different counties within a given MSA. For instance, workers may be willing to buy housing located in a good school district even though it may be far away from their work place. Knowledge spillovers have a positive association with geographical concentration when they are measured at the county level. There are many reasons why knowledge spillovers are significant at a smaller spatial scale. Technological spillovers often require face-to-face contacts, whereas other agglomeration effects such as input-output linkages could take place at a larger scale (Combes and Gobillon, 2015). The geographical scope of knowledge spillovers may be very limited and the county may be a more appropriate geographical scale to capture effects than is the larger MSA.

To control county characteristics, I replace MSA fixed effects with county fixed effects in column 2. Overall, the county-level estimations on Marshallian agglomeration mechanisms generate qualitatively the same results as at the MSA-level estimations in the table 8 but with smaller magnitudes of the estimated effects. The spatial scale of agglomeration effects depends on their type. Whereas industrial customers and labor pooling are vairables having significant impact, knowledge spillovers have a positive impact on spatial concentration at short distances, say within counties.

The negative estimated coefficient on Tax and positive estimated coefficient on $Tax \times EG$ suggest that taxes deter firm births, however, firms in the industries with high EG indices experience relatively low firm births. The result implies that more agglomerated industries are less sensitive to tax differentials. Two proxies for natural advantage are designed to capture local cost advantages. The results indicate that industries with intensive use of coal tend to concentrate in places with rich coal deposits. Low electricity price is more likely to attract firms which heavily rely on electricity in their production. The results in column 1 and column 2 suggest that all sources or mechanisms of agglomeration and local conditions are relevant.

The costs and benefits in firm location choices are often evaluated in the context of agglomeration economies and agglomeration diseconomies (Bhat et al., 2014). Agglomeration economies refer to the benefits that firms experience when locating near one another, while agglomeration diseconomies refer to the negative effects that firms experience when they cluster together. In order to test agglomeration diseconomies, column 3 incorporates county employment which excludes own industry employment. The negative estimated coefficient for county employment implies that more employment may deter firm entries[9]. This result suggests that the crowding effects associated with employment increase (increased rents, congestion costs) offset the benefits of agglomeration (Jofre-Monseny et al., 2011).

[9]When adding the county employment variable to the regression reported in column 3, I note that the coefficients on both tax and county employment variables have surprisingly large magnitudes. I am still investigating the reasons.

Table 8: Estimations of Mfg. Entry Counts. Negative Binomial Regression, County-Industry Level

Dependent Variable:	No. of New Firms		(2004-2011)
	(1)	(2)	(3)
Marshall's factors			
ln Input_{2002}	0.114***	0.019	0.024
	(0.023)	(0.015)	(0.015)
ln Output_{2002}	0.070***	0.050***	0.049**
	(0.033)	(0.019)	(0.019)
ln Labor_{2002}	-0.006	0.010**	0.009*
	(0.012)	(0.005)	(0.005)
ln Citing_{2002}	0.088***	0.024***	0.027***
	(0.013)	(0.007)	(0.007)
ln Cited_{2002}	0.106***	0.003	0.003
	(0.013)	(0.006)	(0.006)
Corporate Tax			
Tax_{2002}	0.003	-0.487***	-4.988***
	(0.069)	(0.017)	(0.803)
Tax$\times$EG index_{2002}	0.106	0.103***	0.093***
	(0.057)	(0.040)	(0.033)
Natural Advantage			
Coal Mining Production_{2002}	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)
Electricity Prices_{2002}	-2.616**	-1.640**	-1.563***
	(1.071)	(0.761)	(0.741)
Control			
County employment_{2002}	N	N	-7.732***
			(1.383)
Own industry employment in county(2002)	Y	Y	Y
Industry fixed effects	Y	Y	Y
County fixed effects	N	Y	Y
MSA fixed effects	Y	N	N
No. of industries	52	52	52
No. of Counties	299	299	299
Observations	15548	15548	15548

Notes: Estimated model: $N_{i,g} = \beta_1 \cdot Marshallian_{i,g} + \beta_2 \cdot T_g + \beta_3 \cdot T_g \cdot EG_i + \beta_4 \cdot Natural\ Advantage_{i,g} + \beta_5 \cdot \log\ emp_{i,g} + \beta_6 \cdot \log\ emp_g + \alpha_i + \alpha_g + \varepsilon_{i,g}$. 299 counties within the top 35 MSAs. Robust standard errors (in parentheses) are clustered at MSA level. $*$ indicates significant at 10%, $**$ indicates significant at 5%, $***$ indicates significant at 1%.

Table 9: Estimations of Mfg. Entry Counts. Marginal Effects at the Means (Negative Binomial Regression), County-Industry Level

Dependent Variable:	No. of New Firms		(2004-2011)
	(1)	(2)	(3)
Marshall's factors			
ln Input_{2002}	0.140***	0.016	0.021
	(0.027)	(0.013)	(0.013)
ln Output_{2002}	0.085***	0.044***	0.043***
	(0.040)	(0.017)	(0.017)
ln Labor_{2002}	−0.007	0.009**	0.008*
	(0.014)	(0.004)	(0.004)
ln Citing_{2002}	0.108***	0.021***	0.023***
	(0.015)	(0.006)	(0.006)
ln Cited_{2002}	0.130***	0.003	0.003
	(0.016)	(0.005)	(0.005)
Corporate Tax			
Tax_{2002}	0.004	−0.426***	−4.294***
	(0.084)	(0.015)	(0.702)
Tax×EG index_{2002}	0.131*	0.090***	0.080***
	(0.071)	(0.035)	(0.029)
Natural Advantage			
Coal Mining Production_{2002}	0.006***	0.004***	0.004***
	(0.002)	(0.001)	(0.001)
Electricity Prices_{2002}	−3.212**	−1.433**	−1.345**
	(1.340)	(0.666)	(0.638)
Control			
County employment_{2002}	N	N	-6.657***
			(1.207)
Own industry employment in county(2002)	Y	Y	Y
Industry fixed effects	Y	Y	Y
County fixed effects	N	Y	Y
MSA fixed effects	Y	N	N
No. of industries	52	52	52
No. of Counties	299	299	299
Observations	15548	15548	15548

Notes: Estimated model: $N_{i,g} = \beta_1 \cdot Marshallian_{i,g} + \beta_2 \cdot T_g + \beta_3 \cdot T_g \cdot EG_i + \beta_4 \cdot Natural\,Advantage_{i,g} + \beta_5 \cdot \log\, emp_{i,g} + \beta_6 \cdot \log\, emp_g + \alpha_i + \alpha_g + \varepsilon_{i,g}$. 299 counties within the top 35 MSAs. Robust standard errors (in parentheses) are clustered at MSA level. $*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

My last topic to discuss is the comparison of the two time periods before and after the Great Recession (sees tables 10-11). Table 10 reports comparisons at the MSA level. Input sharing appears to have a much larger effect on firm births after the financial crisis. In particular, the presence of input suppliers becomes statistically significant. The estimates imply that an increase of 10 employees in industries that supply inputs to industry i creates 1.47 new firms before the recession, and 4.68 new firms in the post-crisis period. In contrast, firm locations appear to be less responsive to local labor market conditions after the recession. An increase of 10 employees in industries that use similar type of workers as those used by industry i increase new firms by 6.22 before the recession, but only 2.53 in the post-crisis period. It is not surprising given the fact that many workers were laid off during the recession, and among them manufacturing workers may have suffered the most. Some workers may change their occupation, or even exit the labor force after long-term unemployment. Knowledge spillovers effects do not seem to be affected by the split of the data. Next, I look at the lower level of aggregation, the county level, in table 11. There has been a strengthening of Marshall forces at the county level. The magnitudes of the estimated coefficients are larger for input-output linkages and labor market pooling. The workings of agglomeration economies have become more pronounced after the Great Recession. One possible explanation is that risk-aversion may play a role in the location choice process. Facing national negative shocks, firms may be more likely to locate in the areas where the clustering of firms may create an advantage to reduce the amount of uncertainty.

Table 10: Comparison of 2004-2007 and 2008-2011, MSA-Industry Level

Dependent Variable: No. of New Firms	(2004-2007)	(2008-2011)
	(1)	(2)
ln Input	0.147	0.448*
	(0.431)	(0.229)
ln Output	0.685***	0.748***
	(0.196)	(0.176)
ln Labor	0.622**	0.253***
	(0.317)	(0.079)
ln Citing	0.013	0.056
	(0.191)	(0.113)
ln Cited	0.219	0.085
	(0.216)	(0.111)
Control		
No. of Estabs in MSA-industry	Y	Y
Industry fixed effects	Y	Y
MSA fixed effects	Y	Y
No. of Industries	52	52
No. of MSA	50	50
Observations	2600	2600

Notes: Includes top 50 MSAs (based on 2000 census population). Robust standard errors (in parentheses) are clustered at MSA level. Independent variables correspond to year 2002 for (2004-2007), and year 2007 for (2008-2011). $*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

Table 11: Comparison of 2004-2007 and 2008-2011, County-Industry Level

Dependent Variable: No. of New Firms	(2004-2007)	(2008-2011)
	(1)	(2)
Marshall's factors		
ln Input	0.006	0.011*
	(0.007)	(0.007)
ln Output	0.020**	0.023***
	(0.009)	(0.006)
ln Labor	0.004*	0.006**
	(0.002)	(0.003)
ln Citing	0.012***	0.007**
	(0.003)	(0.003)
ln Cited	−0.001	0.003
	(0.003)	(0.004)
Corporate Tax		
Tax	−0.181***	−0.157***
	(0.008)	(0.007)
Tax×EG index	0.053**	−0.001
	(0.025)	(0.030)
Natural Advantage		
Coal Mining Production	0.002***	0.003
	(0.001)	(0.002)
Electricity Prices	−0.614*	−0.341
	(0.329)	(0.484)
Control		
Own industry employment in county	Y	Y
Industry fixed effects	Y	Y
County fixed effects	Y	Y
No. of industries	52	52
No. of Counties	299	299
Observations	15548	15548

Notes: 299 counties within the top 35 MSAs. Robust standard errors (in parentheses) are clustered at MSA level. Independent variables correspond to year 2002 for (2004-2007), and year 2007 for (2008-2011). $*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

1.8 Conclusion

This paper contributes to the empirical literature on agglomeration economies and the importance of each of Marshall's agglomeration mechanisms. A unique firm-level data set, the ReferenceUSA Historical Business Database, allows me to explore the determinants of new manufacturing firm locations for the time interval 2004-2011 at different geographic scales.. The richness of the firm-level data set allows me to split the data into pre- and post-crisis time periods, 2004-2007 and 2008-2011. Thus I am able to be one of the first researchers to explore how the workings of agglomeration effects varies before and after the financial crisis.

Considering my findings for the entire time period 2004-2011, my results indicate that proxies for labor market pooling and intermediate good linkages have the most robust effects, positively influencing agglomeration at both the MSA and county levels. Proxies for knowledge spillovers, in contrast, positively affect agglomeration only at the county level. The evidence on input suppliers is somewhat weaker, there appears to be a very limited role for the presence of input suppliers to explain patterns of entry across regions and industries. On the broader level, my paper provides strong support for Marshallian factors relating to labor pooling and input sharing mechanisms, but it does not support the importance of knowledge spillovers. Glaeser and Kerr (2009) also find that the most important mechanism is labor market pooling at the city level during the period of 1976-1999. Similar findings are found to hold for manufacturing sectors in other countries, like Spain. Jofre-Monseny et al. (2011) provide evidence of labor market pooling, followed by input sharing.

It would be interesting to understand whether of not my findings for the manufacturing sector generalize to other sectors, such as services. Many services involve face to face contact which sometimes requires higher transport costs. These services are most likely to concentrate when they can benefit from clustering near customers (Ellison et al., 2010; Glaeser, 2010) . Knowledge spillovers may be more important in innovative sectors, such as those industries located in Silicon Valley.

I do find that some natural advantage variables are very important for new manufacturing firm births. The results suggest that natural advantage can account for a portion of geographic concentration. The role of local taxes in determining the location of new manufacturing firms is identified in the paper. Firm births on average react negatively to corporate taxes but the effects are weaker in the industries that are more geographically concentrated. Overall, these results suggest that local variables do help us to understand the heterogeneity that exists in births of new manufacturing firms.

My last and perhaps most important finding here is that the significance of Marshallian factors does not seem to be affected by the split of the data before and after the Great Recession. However magnitudes do change. There has been a strengthening of local agglomerative forces after the recent chaotic financial times at the national level.New firms may become more risk averse after after large negative shocks and may be more likely to choose a location where industry relations are strong. The presence of agglomerative forces may attract new firms to local areas and therefore help local economies to recover more quickly after recessions. I hope my approach will be useful in future explorations of agglomerative forces for other industrial sectors and in other countries.

References

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2013). The network origins of large economic downturns. *NBER Working Paper 19230*.

Agrawal, A., Cockburn, I., and Rosell, C. (2010). Not invented here? innovation in company towns. *Journal of Urban Economics*, 67(1):78–89.

Agrawal, A., Kapur, D., and McHale, J. (2008). How do spatial and social proximity influence knowledge flows? evidence from patent data. *Journal of Urban Economics*, 64(2):258–269.

Arauzo-Carod, J.-M. (2008). Industrial location at a local level: comments on the territorial level of the analysis. *Tijdschrift voor Economische en Sociale Geografie*, 99(2):193–208.

Arauzo-Carod, J.-M., Liviano-Solis, D., and Manjón-Antolín, M. (2010). Empirical studies in industrial location: An assessment of their methods and results. *Journal of Regional Science*, 50(3):685–711.

Autant-Bernard, C. (2006). Where do firms choose to locate their r&d? a spatial conditional logit analysis on french data. *European Planning Studies*, 14(9):1187–1208.

Bartik, T. J. (1985). Business location decisions in the united states: Estimates of the effects of unionization, taxes, and other characteristics of states. *Journal of Business and Economic Statistics*, 3(1):14–22.

Bhat, C. R., Paleti, R., and Singh, P. (2014). A spatial multivariate count model for firm location decisions. *Journal of Regional Science*, 54(3):462–502.

Brülhart, M., Jametti, M., and Schmidheiny, K. (2012). Do agglomeration economies reduce the sensitivity of firm location to tax differentials? *The Economic Journal*, 122(563):1069–1093.

Carlton, D. (1983). The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *The Review of Economics and Statistics*, 65(3):440–449.

Chauvin, J. P., Glaeser, E. L., and Tobio, K. (2013). Urban economics in the us and india. *Manuscript in progress, Harvard University*.

Combes, P.-P., Démurger, S., and Li, S. (2015). Migration externalities in chinese cities. *European Economic Review*, 76:152–167.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. *Handbook of Regional and Urban Economics*, 5:247–348.

Delgado, M., Porter, M. E., and Stern, S. (2010). Clusters and entrepreneurship. *Journal of Economic Geography*, 10(4):495–518.

Delgado, M., Porter, M. E., and Stern, S. (2015). Clusters and the great recession. Technical report.

Diamond, C. A. and Simon, C. J. (1990). Industrial specialization and the returns to labor. *Journal of Labor Economics*, pages 175–201.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217.

Duranton, G. (2015). Growing through cities in developing countries. *World Bank Research Observer*, 30(1).

Duranton, G. (2016). Agglomeration effects in colombia. *Journal of Regional Science*, 56(2):210–238.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. *Handbook of regional and urban economics*, 4:2063–2117.

Egeln, J., Gottschalk, S., and Rammer, C. (2004). Location decisions of spin-offs from public research institutions. *Industry and Innovation*, 11(3):207–223.

Ellison, G. and Glaeser, E. L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration? *The American Economic Review*, 89(2):311–316.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *The American Economic Review*, 100(3):1195–1213.

Glaeser, E. L. (2010). Introduction to the economics of agglomeration. *Agglomeration Economics*, pages 1–14. University of Chicago Press.

Glaeser, E. L. and Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain? *Journal of Economics and Management Strategy*, 18(3):623–663.

Glaeser, E. L. and Kohlhase, J. E. (2004). Cities, regions and the decline of transport costs. *Papers in Regional Science*, 83(1):197–228.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Guimaraes, P., Figueirdo, O., and Woodward, D. (2003). A tractable approach to the firm location decision problem. *Review of Economics and Statistics*, 85(1):201–204.

Guimaraes, P., Figueiredo, O., and Woodward, D. (2000). Agglomeration and the location of foreign direct investment in portugal. *Journal of Urban Economics*, 47(1):115–135.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–598.

Jofre-Monseny, J., Marín-López, R., and Viladecans-Marsal, E. (2011). The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics*, 70(2):61–74.

Jofre-Monseny, J. and Solé-Ollé, A. (2012). Which communities should be afraid of mobility? the effects of agglomeration economies on the sensitivity of employment location to local taxes. *Regional Science and Urban Economics*, 42(1):257–268.

Kim, S. (1999). Regions, resources, and economic geography: Sources of us regional comparative advantage, 1880–1987. *Regional Science and Urban Economics*, 29(1):1–32.

Marshall, A. (1890). Principles of economics. Macmillan London.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142. Academic Press.

Overman, H. G. and Puga, D. (2002). Unemployment clusters across europe's regions and countries. *Economic Policy*, 17(34):115–148.

Porter, M. E. (1990). The competitive advantage of nations. *Harvard Business Review*.

Porter, M. E. (2007). Clusters and economic policy: Aligning public policy with the new economics of competition.

Roberto, B. (2004). Acquisition versus greenfield investment: the location of foreign manufacturers in italy. *Regional Science and Urban Economics*, 34(1):3–25.

Rocha, N. (2008). Firm location determinants: Empirical evidence for france. HEID Working Paper No. 08. Graduate Institute of International and Development Studies.

Rohlin, S., Rosenthal, S. S., and Ross, A. (2014). Tax avoidance and business location in a state border model. *Journal of Urban Economics*, 83:34–49.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85(2):377–393.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics*, 4:2119–2171.

Wu, F. (1999). Intrametropolitan fdi firm location in guangzhou, china a poisson and negative binomial analysis. *The Annals of Regional Science*, 33(4):535–555.

Appendix A: Tobit Model

Table 12: Marginal Effects at the Means (Tobit Model), MSA-Industry Level

Dependent Variable	No. of New Firms (2004-2011)			
	(1)	(2)	(3)	(4)
ln $Input_{2002}$	3.325			4.646
	(5.079)			(5.555)
ln $Output_{2002}$	8.340**			9.358**
	(3.517)			(3.701)
ln $Labor_{2002}$		1.458		-4.900
		(3.105)		(3.855)
ln $Citing_{2002}$			4.514	4.833
			(3.899)	(4.027)
ln $Cited_{2002}$			-0.511	-1.282
			(4.097)	(4.107)
Control				
Industry fixed effects	Y	Y	Y	Y
MSA fixed effects	Y	Y	Y	Y
No. of Industries	52	52	52	52
No. of MSA	50	50	50	50
Observations	2600	2600	2600	2600

Notes: Includes top 50 MSAs (based on 2000 census population). $*$ indicates significant at 10%, $**$ indicates significant at 5%, $***$ indicates significant at 1%.

Table 13: Marginal Effects at the Means (Tobit Model), County-Industry Level

Dependent Variable:	No. of New Firms	(2004-2011)			
	(1)	(2)	(3)	(4)	(5)
Marshall's factors					
ln Input_{2007}	0.069			-0.070	-0.077
	(0.209)			(0.216)	(0.216)
ln Output_{2007}	0.159			0.032	0.035
	(0.149)			(0.153)	(0.153)
ln Labor_{2007}		0.218***		0.175*	0.171*
		(0.085)		(0.090)	(0.090)
ln Citing_{2007}			0.153	0.127	0.128
			(0.100)	(0.101)	(0.101)
ln Cited_{2007}			0.146	0.131	0.130
			(0.101)	(0.101)	(0.101)
Corporate Tax					
Tax_{2007}	N	N	N	N	-4.318***
					(0.533)
Tax×EG index_{2007}	N	N	N	N	0.861
					(0.912)
Natural Advantage					
Coal Mining Production_{2007}	N	N	N	N	0.061
					(0.057)
Electricity Prices_{2007}	N	N	N	N	-14.302*
					(7.389)
Control					
Industry fixed effects	Y	Y	Y	Y	Y
County fixed effects	Y	Y	Y	Y	Y
No. of industries	52	52	52	52	52
No. of Counties	299	299	299	299	299
Observations	15548	15548	15548	15548	15548

Notes: 299 counties within the top 35 MSAs. $*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

Appendix B: Normalized Dependent Variables

Table 14: Marginal Effects at the Means (Negative Binomial Regression), MSA-Industry Level

Dependent Variable	$\frac{No.of\ New\ Firms}{Land\,area}$ (2004-2011)			
	(1)	(2)	(3)	(4)
ln $Input_{2002}$	0.0006***			0.0005**
	(0.0002)			(0.0002)
ln $Output_{2002}$	0.0005***			0.0003***
	(0.0001)			(0.0001)
ln $Labor_{2002}$		0.0006***		0.0005***
		(0.0001)		(0.00001)
ln $Citing_{2002}$			0.0001*	-0.0001
			(0.0001)	(0.0001)
ln $Cited_{2002}$			0.0001	0.0001
			(0.0001)	(0.0001)
Control				
Industry fixed effects	Y	Y	Y	Y
MSA fixed effects	Y	Y	Y	Y
No. of Industries	52	52	52	52
No. of MSA	50	50	50	50
Observations	2600	2600	2600	2600

Notes: Includes top 50 MSAs (based on 2000 census population). Robust standard errors (in parentheses) are clustered at MSA level. $*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

Table 15: Marginal Effects at the Means (Negative Binomial Regression), County-Industry Level

Dependent Variable	$\frac{No.of\, New\, Firms}{Land\, area}$		(2004-2011)	
	(1)	(2)	(3)	(4)
ln Input_{2002}	−0.00007			−0.00010*
	(0.00007)			(0.00005)
ln Output_{2002}	0.00004			0.00004
	(0.00011)			(0.00009)
ln Labor_{2002}		0.00007***		0.00008***
		(0.00003)		(0.00002)
ln Citing_{2002}			0.00010***	0.00010***
			(0.00002)	(0.00002)
ln Cited_{2002}			−0.00003	−0.00004
			(0.00001)	(0.00002)
Control				
Industry fixed effects	Y	Y	Y	Y
County fixed effects	Y	Y	Y	Y
No. of Industries	52	52	52	52
No. of MSA	299	299	299	299
Observations	15548	15548	15548	15548

Notes: 299 counties within the top 35 MSAs. Robust standard errors (in parentheses) are clustered at MSA level. $*$ indicates significant at 10%, $**$ indicates significant at 5%,$***$ indicates significant at 1%.

Appendix C: MSAs List (In Descending Order by Population)[10]

1. New York-Newark-Edison, NY-NJ-PA Metro Area
2. Los Angeles-Long Beach-Santa Ana, CA Metro Area
3. Chicago-Naperville-Joliet, IL-IN-WI Metro Area
4. Dallas-Fort Worth-Arlington, TX Metro Area
5. Houston-Baytown-Sugar Land, TX Metro Area
6. Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area
7. Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area
8. Miami-Fort Lauderdale-Miami Beach, FL Metro Area
9. Atlanta-Sandy Springs-Marietta, GA Metro Area
10. Boston-Cambridge-Quincy, MA-NH Metro Area
11. San Francisco-Oakland-Fremont, CA Metro Area
12. Phoenix-Mesa-Scottsdale, AZ Metro Area
13. Riverside-San Bernardino-Ontario, CA Metro Area
14. Detroit-Warren-Livonia, MI Metro Area
15. Seattle-Tacoma-Bellevue, WA Metro Area
16.Minneapolis-St. Paul-Bloomington, MN-WI Metro Area
17. San Diego-Carlsbad-San Marcos, CA Metro Area
18. Tampa-St. Petersburg-Clearwater, FL Metro Area
19. Denver-Aurora, CO Metro Area
20. St. Louis, MO-IL Metro Area
21. Baltimore-Towson, MD Metro Area
22. Charlotte-Gastonia-Concord, NC-SC Metro Area
23. Portland-Vancouver-Beaverton, OR-WA Metro Area
24. Orlando, FL Metro Area
25. San Antonio, TX Metro Area
26. Pittsburgh, PA Metro Area
27. Sacramento–Arden-Arcade–Roseville, CA Metro Area
28. Cincinnati-Middletown, OH-KY-IN Metro Area
29. Las Vegas-Paradise, NV Metro Area
30. Kansas City, MO-KS Metro Area
31. Cleveland-Elyria-Mentor, OH Metro Area
32. Columbus, OH Metro Area
33. Austin-Round Rock, TX Metro Area
34. Indianapolis, IN Metro Area

[10]Office of Management and Budget (OMB) 2003 Delineartion http://www.census.gov/population/metro/data/defhist.html

35. San Jose-Sunnyvale-Santa Clara, CA Metro Area
36. Nashville-Davidson–Murfreesboro, TN Metro Area
37. Virginia Beach-Norfolk-Newport News, VA-NC Metro Area
38. Providence-New Bedford-Fall River, RI-MA Metro Area
39. Milwaukee-Waukesha-West Allis, WI Metro Area
40. Jacksonville, FL Metro Area
41. Oklahoma City, OK Metro Area
42. Memphis, TN-MS-AR Metro Area
43. Louisville, KY-IN Metro Area
44. Raleigh-Cary, NC Metro Area
45. Richmond, VA Metro Area
46. New Orleans-Metairie-Kenner, LA Metro Area
47. Hartford-West Hartford-East Hartford, CT Metro Area
48. Salt Lake City, UT Metro Area
49. Birmingham-Hoover, AL Metro Area
50. Buffalo-Cheektowaga-Tonawanda, NY Metro Area

Appendix D: County Components [11]

1. New York-Newark-Edison, NY-NJ-PA Metro Area

   – Bergen, NJ Essex, NJ Hudson, NJ Hunterdon, NJ Middlesex, NJ Monmouth, NJ Morris, NJ Ocean, NJ Passaic, NJ Somerset, NJ Sussex, NJ Union, NJ Bronx, NY Kings, NY Nassau, NY New York, NY Putnam, NY Queens, NY Richmond, NY Rockland, NY Suffolk, NY Westchester, NY Pike, PA

2. Los Angeles-Long Beach-Santa Ana, CA Metro Area

   – Los Angeles, CA Orange, CA

3. Chicago-Naperville-Joliet, IL-IN-WI Metro Area

   – Cook, IL DeKalb, IL DuPage, IL Grundy, IL Kane, IL Kendall, IL Lake, IL McHenry, IL Will, IL Jasper, IN Lake, IN Newton, IN Porter, IN Kenosha, WI

4. Dallas-Fort Worth-Arlington, TX Metro Area

   – Collin, TX Dallas, TX Delta, TX Denton, TX Ellis, TX Hunt, TX Johnson, TX Kaufman, TX Parker, TX Rockwall, TX Tarrant, TX Wise, TX

5. Houston-Baytown-Sugar Land, TX Metro Area

   – Austin, TX Brazoria, TX Chambers, TX Fort Bend, TX Galveston, TX Harris, TX Liberty, TX Montgomery, TX San Jacinto, TX Waller, TX

6. Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area

   – District of Columbia, DC Calvert, MD Charles, MD Frederick, MD Montgomery, MD Prince George's, MD Alexandria city, VA Arlington, VA Clarke, VA Fairfax, VA Fairfax city, VA Falls Church city, VA Fauquier, VA Fredericksburg city, VA Loudoun, VA Manassas city, VA Manassas Park city, VA Prince William, VA Spotsylvania, VA Stafford, VA Warren, VA Jefferson, WV

7. Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area

   – New Castle, DE Cecil, MD Burlington, NJ Camden, NJ Gloucester, NJ Salem, NJ Bucks, PA Chester, PA Delaware, PA Montgomery, PA Philadelphia, PA

8. Miami-Fort Lauderdale-Miami Beach, FL Metro Area

   – Broward, FL Miami-Dade, FL Palm Beach, FL

[11] 50 largest MSAs using OMB 2003 delineations. Component counties given by county name, state.

9. Atlanta-Sandy Springs-Marietta, GA Metro Area

– Barrow, GA Bartow, GA Butts, GA Carroll, GA Cherokee, GA Clayton, GA Cobb, GA Coweta, GA Dawson, GA DeKalb, GA Douglas, GA Fayette, GA Forsyth, GA Fulton, GA Gwinnett, GA Haralson, GA Heard, GA Henry, GA Jasper, GA Lamar, GA Meriwether, GA Newton, GA Paulding, GA Pickens, GA Pike, GA Rockdale, GA Spalding, GA Walton, GA

10. Boston-Cambridge-Quincy, MA-NH Metro Area

– Essex, MA Middlesex, MA Norfolk, MA Plymouth, MA Suffolk, MA Rockingham, NH Strafford, NH

11. San Francisco-Oakland-Fremont, CA Metro Area

– Alameda, CA Contra Costa, CA Marin, CA San Francisco, CA San Mateo, CA

12. Phoenix-Mesa-Scottsdale, AZ Metro Area

– Maricopa, AZ Pinal, AZ

13. Riverside-San Bernardino-Ontario, CA Metro Area

– Riverside, CA San Bernardino, CA

14. Detroit-Warren-Livonia, MI Metro Area

– Lapeer, MI Livingston, MI Macomb, MI Oakland, MI St. Clair, MI Wayne, MI

15. Seattle-Tacoma-Bellevue, WA Metro Area

– King, WA Pierce, WA Snohomish, WA

16. Minneapolis-St. Paul-Bloomington, MN-WI Metro Area

– Anoka, MN Carver, MN Chisago, MN Dakota, MN Hennepin, MN Isanti, MN Ramsey, MN Scott, MN Sherburne, MN Washington, MN Wright, MN Pierce, WI St. Croix, WI

17. San Diego-Carlsbad-San Marcos, CA Metro Area

– San Diego, CA

18. Tampa-St. Petersburg-Clearwater, FL Metro Area

– Hernando, FL Hillsborough, FL Pasco, FL Pinellas, FL

19. Denver-Aurora, CO Metro Area

- Adams, CO   Arapahoe, CO   Broomfield, CO   Clear Creek, CO   Denver, CO   Douglas, CO   Elbert, CO   Gilpin, CO   Jefferson, CO   Park, CO

20. St. Louis, MO-IL Metro Area
    - Bond, IL   Calhoun, IL   Clinton, IL   Jersey, IL   Macoupin, IL   Madison, IL   Monroe, IL   St. Clair, IL   Crawford, MO   Franklin, MO   Jefferson, MO   Lincoln, MO   St. Charles, MO   St. Louis, MO   St. Louis city, MO   Warren, MO   Washington, MO
21. Baltimore-Towson, MD Metro Area
    - Anne Arundel, MD   Baltimore, MD   Baltimore city, MD   Carroll, MD   Harford, MD   Howard, MD   Queen Anne's, MD
22. Charlotte-Gastonia-Concord, NC-SC Metro Area
    - Anson, NC   Cabarrus, NC   Gaston, NC   Mecklenburg, NC   Union, NC   York, SC
23. Portland-Vancouver-Beaverton, OR-WA Metro Area
    - Clackamas, OR   Columbia, OR   Multnomah, OR   Washington, OR   Yamhill, OR   Clark, WA   Skamania, WA
24. Orlando, FL Metro Area
    - Lake, FL   Orange, FL   Osceola, FL   Seminole, FL
25. San Antonio, TX Metro Area
    - Atascosa, TX   Bandera, TX   Bexar, TX   Comal, TX   Guadalupe, TX   Kendall, TX   Medina, TX   Wilson, TX
26. Pittsburgh, PA Metro Area
    - Allegheny, PA   Armstrong, PA   Beaver, PA   Butler, PA   Fayette, PA   Washington, PA   Westmoreland, PA
27. Sacramento–Arden-Arcade–Roseville, CA Metro Area
    - El Dorado, CA   Placer, CA   Sacramento, CA   Yolo, CA
28. Cincinnati-Middletown, OH-KY-IN Metro Area
    - Dearborn, IN   Franklin, IN   Ohio, IN   Boone, KY   Bracken, KY   Campbell, KY   Gallatin, KY   Grant, KY   Kenton, KY   Pendleton, KY   Brown, OH   Butler, OH   Clermont, OH   Hamilton, OH   Warren, OH
29. Las Vegas-Paradise, NV Metro Area

- Clark, NV

30. Kansas City, MO-KS Metro Area
    - Franklin, KS Johnson, KS Leavenworth, KS Linn , KS Miami, KS Wyandotte, KS Bates, MO Caldwell, MO Cass, MO Clay, MO Clinton, MO Jackson, MO Lafayette, MO Platte, MO Ray, MO
31. Cleveland-Elyria-Mentor, OH Metro Area
    - Cuyahoga, OH Geauga, OH Lake, OH Lorain, OH Medina, OH
32. Columbus, OH Metro Area
    - Delaware, OH Fairfield, OH Franklin, OH Licking, OH Madison, OH Morrow, OH Pickaway, OH Union, OH
33. Austin-Round Rock, TX Metro Area
    - Bastrop, TX Caldwell, TX Hays, TX Travis, TX Williamson, TX
34. Indianapolis, IN Metro Area
    - Boone, IN Brown, IN Hamilton, IN Hancock, IN Hendricks, IN Johnson, IN Marion, IN Morgan, IN Putnam, IN Shelby, IN
35. San Jose-Sunnyvale-Santa Clara, CA Metro Area
    - San Benito, CA Santa Clara, CA
36. Nashville-Davidson–Murfreesboro, TN Metro Area
    - Cannon, TN Cheatham, TN Davidson, TN Dickson, TN Hickman, TN Macon, TN Robertson, TN Rutherford, TN Smith, TN Sumner, TN Trousdale, TN Williamson, TN Wilson, TN
37. Virginia Beach-Norfolk-Newport News, VA-NC Metro Area
    - Currituck, NC Chesapeake city, VA Gloucester, VA Hampton city, VA Isle of Wight, VA James City, VA Mathews, VA Newport News city, VA Norfolk city, VA Poquoson city, VA Portsmouth city, VA Suffolk city, VA Surry, VA Virginia Beach city, VA Williamsburg city, VA York, VA
38. Providence-New Bedford-Fall River, RI-MA Metro Area
    - Bristol, MA Bristol, RI Kent, RI Newport, RI Providence, RI Washington, RI
39. Milwaukee-Waukesha-West Allis, WI Metro Area

- Milwaukee, WI   Ozaukee, WI   Washington, WI   Waukesha, WI

40. Jacksonville, FL Metro Area

  - Baker, FL   Clay, FL   Duval, FL   Nassau, FL   St. Johns, FL

41. Oklahoma City, OK Metro Area

  - Canadian, OK   Cleveland, OK   Grady, OK   Lincoln, OK   Logan, OK   McClain, OK   Oklahoma, OK

42. Memphis, TN-MS-AR Metro Area

  - Crittenden, AR   DeSoto, MS   Marshall, MS   Tate, MS   Tunica, MS   Fayette, TN   Shelby, TN   Tipton, TN

43. Louisville, KY-IN Metro Area

  - Clark, IN   Floyd, IN   Harrison, IN   Washington, IN   Bullitt, KY   Henry, KY   Jefferson, KY   Meade, KY   Nelson, KY   Oldham, KY   Shelby, KY   Spencer, KY   Trimble, KY

44. Raleigh-Cary, NC Metro Area

  - Franklin, NC   Johnston, NC   Wake , NC

45. Richmond, VA Metro Area

  - Amelia, VA   Caroline, VA   Charles City, VA   Chesterfield, VA   Colonial Heights city, VA   Cumberland, VA   Dinwiddie, VA   Goochland, VA   Hanover, VA   Henrico, VA   Hopewell city, VA   King and Queen, VA   King William, VA   Louisa, VA   New Kent, VA   Petersburg city, VA   Powhatan, VA   Prince George, VA   Richmond city, VA   Sussex, VA

46. New Orleans-Metairie-Kenner, LA Metro Area

  - Jefferson Parish, LA   Orleans Parish, LA   Plaquemines Parish, LA   St. Bernard Parish, LA   St. Charles Parish, LA   St. John the Baptist Parish, LA   St. Tammany Parish, LA

47. Hartford-West Hartford-East Hartford, CT Metro Area

  - Hartford, CT   Middlesex, CT   Tolland, CT

48. Salt Lake City, UT Metro Area

  - Salt Lake, UT   Summit, UT   Tooele, UT

49. Birmingham-Hoover, AL Metro Area

  - Bibb, AL   Blount, AL   Chilton, AL   Jefferson, AL   Shelby, AL   St. Clair, AL   Walker, AL

50. Buffalo-Cheektowaga-Tonawanda, NY Metro Area

- Erie, NY   Niagara, NY

Chapter 2

A Note on Geographical Constraints and Housing Markets in China[1]

2.1 Introduction

Real estate markets in China have experienced extremely fast growth since 2000 with dramatic increases in prices in some areas. There are many factors that may have led to rising housing prices. Demographic variables, such as population size and its growth are important determinants of the demand, and other factors, such as income level, and the cost of credit, play critical roles in housing markets. In recent years, empirical studies on this topic have shifted focus to the role that housing supply has played in explaining the variation in housing prices and their changes in housing markets. In this paper, I ask whether geographic land constraints and government land regulation contribute to price and quantity changes in Chinese housing markets.

An important research question is whether supply constraints are related to price movements. Housing supply elasticities have received much attention in the past decade (Quigley and Raphael, 2005; Glaeser, Gyourko, and Sakes 2005; Gyourko and Saiz, 2006). Evidence that geography and land use regulations affect housing supply elasticity is now widely documented for U.S. cities (Gyourko, Saiz, and Summers, 2008; Glaeser, Gyourko, and Saiz, 2008; Mian and Sufi, 2009; Saiz, 2010). However, some authors argue that the elasticity of supply does not affect housing market dynamics during bubble periods (Davidoff, 2013). Empirical studies have yielded mixed results.

In urban China, the bulk of land is under the control of government officials and becomes available for housing markets only via decisions made by government. The existing literature on this topic has focused on the determinants and potential impacts of land policy on housing supply elasticities (Zheng and Kahn, 2008; Cao, Feng, and Tao, 2008; Fu, Zheng, and Liu, 2009; Du, Ma, and An, 2011; Fu, Zheng, and Liu, 2012).

[11] Chapter 2 is now published in the *Journal of Housing Economics* (Dong, 2016). My paper appears here with permission from the journal publisher that the full paper can be included in a dissertation for non-commercial purposes. https://www.elsevier.com/about/our-business/policies/copyright/permissions

Recent contributions to the literature on land value show that house price growth is driven by increasing land values and that there is heterogeneity in land price across cities (Peng and Thibodeau, 2012; Deng, Gyourko, and Wu, 2012; Wu, Gyourko, and Deng, 2015). To the best of my knowledge, Wang, Chan, and Xu (2012) is the only paper studying geography and Chinese housing (hereafter WCX (2012). They find that geographic, economic and regulatory factors (zoning rules and government revenues) determine housing supply elasticities across cities.

I extend the work of WCX (2012) in at least two ways by integrating important aspects of the Chinese land quota system into my model. First I examine whether land supply via government decisions is quasi-exogenous to price growth and second whether or not government decisions change the impact of geographical constraints on housing prices. An upward sloping supply curve can possibly exist in the market where land is discretely allocated by the government, such as China, when incorporating the factor of natural land constraints. Moreover, my calculation method to define geographical land constraints differs from the methodology used by WCX (2012)[2].

Based on my model estimated for a sample of 35 cities in China from 2003 to 2012, I find that cities with less naturally-available land have experienced greater price appreciation and smaller quantity response. Moreover, I find that the allocations of land use via government decisions is unresponsive to changes in housing price and quantity, suggesting decisions of governments about land supply allocations may not be dependent on housing prices.

The structure of the paper is as follows: Section 2 discusses the workings of the Chinese land quota system. Section 3 discusses the possibility of upward sloping land supply in Chinese urban markets. Section 4 describes the data and Section 5 explains the empirical estimation strategy. Section 6 presents the results.

2.2 China land quota system

[2] The city radius their paper uses to calculate the developable land ratio is three times the conceptual city radius (the radius that makes a circle have a similar area as an urban built-up area). The average real radius for the 35 cities is 30.50 kilometers. In contrast, I use a predetermined measure (invariant radius, 35 kilometers for Chinese cities), which is exogenous to demand conditions and the level of development that city has had.

China has experienced rapid urbanization over the past decade, which has resulted in a large amount of arable land being used for non-agriculture purposes. In the face of a massive loss of arable land, the central government imposed a quota system, which is used to control land use and conversion of farmland into urban land (Lichtenberg and Ding, 2008). The land quota system includes three main parts: the maximum amount of land to be used for construction, the minimum amount of farmland to be maintained, and an annual quota for the amount of newly added construction land that is transferred from farmland (Lichtenberg and Ding, 2009; Tan et al. 2011; Xiao and Zhao, 2015). This land quota system is planned at the central level. The central government is in charge of approving the conversion of farmland to construction land use and the supply of newly added construction land. Local governments are responsible for putting land on the market through open auction and are given the autonomy to decide the purpose the land is to be used for (manufacturing, commercial and residential) (Cai, 2011). Therefore, land supply in China can be considered as quasi-exogenous (Liu and Huang, 2016; Peng and Thibodeau, 2011). Given data limitations, I mainly focus on the third part of the quota system, newly-added construction land.

2.3 Theory: Possibility of upward sloping supply curve

It is useful to discuss whether geography matters in the market where land is discretely allocated by the government, such as China. Empirical studies on the topic of urban land use conducted to date have been based on the Ricardian theory of rent and the monocentric model. In the model, the friction of distance generates a rent gradient between the city center and the periphery. As cities expand in population, large amounts of land are set aside for new development and this land is located further away from the city center. As a result, the rent gradient shifts upwards and becomes steeper (Wheaton, 2004). Making use of this theory can explain why there may be an upward slope to land supply in Chinese urban markets. Even if the land supply is exogenously-driven by discretionary policies, natural constraints still matter in Chinese urban markets. The more constrained the city is, the further from the center the land has to be developed. Positive demand shocks should imply increasing Ricardian rents on land throughout the city[3].

2.4 Data

This paper examines the relations between geography, government regulation and housing markets in China. I consider 35 major cities in China: 4 municipalities which are under the Chinese central government's direct administration and the capital cities of 31 provinces. Annual data on real housing prices (from 2003 to 2012) and newly-built apartment units from (2005 to 2012) are retrieved from China Real Estate Yearbook for various years[4].

I mainly focus on two proxy variables for the supply-side of housing. I start with geographical constraints. My work on geographical constraints is inspired by Saiz (2008 and 2010) and is adjusted to the Chinese data. Saiz (2008 and 2010) provides a measure of exogenously undevelopable land in cities in the US. He estimates the proportion of land unsuitable for housing development by using GIS software and an elevation model. He assumes that it is too costly to develop terrains with high slopes and infill the areas with water. Therefore, the calculation excludes the share of land in a 50-kilometer radius around the geographic centroid of a metropolitan area (Saiz, 2008 and 2010) that has a slope of larger than 15 degrees and land lost to bodies of water. The USGS Digital Elevation Model (DEM) that Saiz used is based on 90 square meter cell grids (cells 90 meters wide by 90 meters long, (Saiz, 2008 and 2010)).

For the Chinese housing markets, I use digital mapping data from the ASTER Global Digital Elevation Model, 30 meter resolution, version2 (updated in 2011) which is jointly produced by the NASA/Japan ASTER team. The data is available on the Japanese Spatial System[5], and is used to calculate how much of the land around each city exhibits slopes above 15% and the area lost to oceans. For the area lost to internal water, I then use a

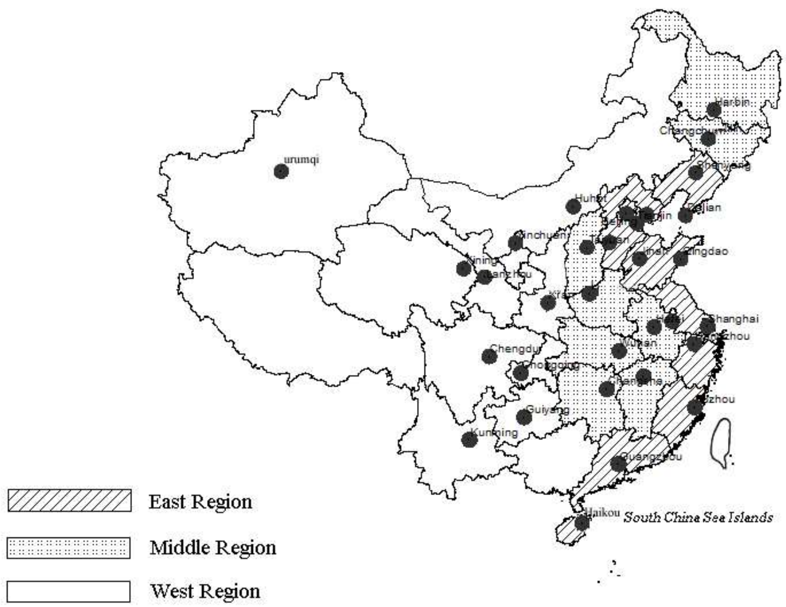
---

[3] I would like to thank Dr. Albert Saiz for his discussion of my paper at the 4th International Workshop on Regional, Urban, and Spatial Economics Conference at Tsinghua University, June 2015. Section 3 is based on Dr Saiz's comments at the conference.

[4] Real housing prices are measured as average selling price (yuan/square meter) for all new homes, calculated by dividing total sales value by total floor space at the city level. Both housing price and completed units are reported in China Real Estate Yearbook, chapter 7, comprehensive real estate data for 35 cities.

[5] Data Source: http://gdem.ersdac.jspacesystems.or.jp/search.jsp

shapefile of water bodies provided by the City University of New York, International ESRI Data[6]. My paper uses a predetermined measure to define city radius, an invariant radius of 35 kilometers[7] for Chinese cities, which is exogenous to demand conditions and the level of development that the city has had. Appendix Table A displays the data and ranks the cities based on the calculated developable land share. The mean of developable land share is 73.5%, with standard deviation 17.6%. Figure 1 exhibits 35 cities distributed across the nation and table 1 displays the percentages of undevelopable area for 35 cities for which housing prices are publicly available.



Source: Author's own GIS drawing, city point shapefile provided by City University of New York, International ESRI Data. 60-kilometers buffers created around cities centers for clear visual effect.

Figure 1: Distribution of 35 Cities across China

[6] Data Source: Baruch Geoportal:International,http://www.baruch.cuny.edu/geoportal/data/esri/esri-intl.htm

[7] The definition of urban area in China is significantly different from classifications used in North America. City (Prefectural) level is the second level of the administrative structure following provincial level (first level) in China. The boundaries of Chinese cities consist of an urbanized core surrounded by towns and rural regions. Therefore Chinese urban administrative area are different from what can be called the urban statistical area, urban areas in China generally refer to the urbanized core within city(Chan, 2013).Distances to work are generally much shorter in China since the majority of urban residents rely on public transportation and bicycles (Leman, 2005).

Table 1: Chinese Cities Developable Land Shares (%)

Northeast area		Southeast area		Middle area		West area	
Shenyang	93.6	Hefei	92.0	Zhengzhou	93.3	Yinchuan	87.5
Tianjin	92.8	Guangzhou	87.6	Changsha	88.4	Chengdu	85.0
Harbin	92.7	Shanghai	81.4	Nanchang	87.7	Xi'an	83.0
Shijiazhuang	92.4	Nanjing	80.6	Wuhan	81.3	Urumqi	78.3
Changchun	91.4	Hangzhou	78.4	Taiyuan	65.7	Huhehot	75.8
Beijing	87.1	Haikou	58.7			Nanning	73.2
Jinan	81.1	Ningbo	56.1			Chongqing	67.5
Qingdao	45.7	Fuzhou	49.2			Guiyang	63.8
Dalian	33.7	Xiamen	48.9			Kunming	53.2
						Xining	58.1
						Lanzhou	45.6

Notes: Developable land share represents the share of land that is suitable for housing construction, and is derived from the ASTER Global Digital Elevation Model (30m resolution, version 2, Oct. 2011) and the author's own GIS calculations. Internal water shapefile provided by City University of New York, International ESRI Data.

One concern is whether the calculation of developable land share is able to capture urban growth. The elevation model only reflects natural geographic characteristics (height of land takes on positive values, elevation of sea level is set to 0 and below sea level, the elevations are negative values) rather than man-made things. Unless there is a natural disaster, such as an earthquake, that changes the height of land or the location of the water, the developable land share is constant over time. The measure does not vary over time with urban growth or man-made construction. The measure is completely exogenous by its nature.

Man-made constraints are other important proxy variables for housing supply. A large number of empirical studies exploring the sources of variation in housing supply elasticities in the US focus on regulatory policies on land and residential construction. Chinese housing markets have faced similar situations concerning regulations of the housing and land markets, with more restraints on land supply. As mentioned earlier, I focus on the newly-added construction-use land, the use right of which has been granted though the open auction in the land market. Newly-added land area refers to the area of farmland and unused land transferred and requisitioned after approval according to law. Farmland is the main source of land for urban expansion. The rapid rate of urban land expansion causes concern because of potential threats to farmland. The newly- added land supply for construction-use has mirrored the government policy on land supply. The data has been consistently recorded since 2003 at the city level. Therefore, I focus on the housing and land market covering the period 2003-2012.

Demand controls, such as population and income are obtained from China City Statistical Year book, China City Construction Statistical Yearbook and China population census in 2000, 2010. Appendix Table B displays summary statistics for these variables.

2.5 Empirical specification

The objectives of the paper are to study the relation between price (and quantity) movement and proxies for supply across cities, and to assess the importance of geographic constraints and government land regulation in explaining local housing supply responses. I start by examining the long changes of price (2003-2012) and quantity (2005-2012) associated with my two supply side proxies.

Theoretically, price and quantity of housing are determined by the interaction of demand and supply. In addition to price, demand is assumed to be a function of variables reflecting market activities (such as demographics and income) and supply is assumed to be a function of variables including the land supply stock (Wang, Tu, and Li, 2015). The equilibrium conditions could be represented by reduced-form expressions; I do not estimate structural models here. My motivation is as follows. Assuming a linear first-order approximation to the housing supply function, the overall change in price in the city is given by demand shocks D interacted with the supply parameterβ: $\frac{P_t}{P_0} = \beta \cdot D$. Then, taking the log to both sides, I obtain:

$$\Delta \log P = \log \beta + \log D, \tag{1}$$

where demand shocks D is a function of social and economic fundamentals (x), and the supply parameter β is a function of geographic and man-made land constraints as follows:

$$D = f(x) \tag{2}$$

and

$$\beta = \exp(developable\ land\ share + newly\ increased\ land). \tag{3}$$

Plugging equation (2) and (3) back into (1), I get:

$$\Delta \log P = devshr + newly\ increased\ land + \log f(x). \tag{4}$$

Throughout this paper, I focus on the long-run housing dynamics rather than high-frequency volatility. Thus, a cross-sectional study allows for comparisons of long changes across regions. The regression model can be expressed as:

$$\Delta \log Price_{\mathrm{i,03-12}} = \alpha_0 + \alpha_1 \cdot devshr_i + \alpha_2 Newly\ increased\ land_i + \alpha_3 \cdot X_i + u_i. \tag{5}$$

where real housing price is measured in yuan/square meters. $Newly\ increaseded\ land_i$[8] is proxy to capture the government land policy (man-made regulation), $devshr_i$ denotes naturally land available for development (geography constraints). The developable land share is assumed to be uncorrelated with unobservable drivers of demand after controlling for X_i , whereas X_i is a vector of control variables[9]. Conditional on demand shocks, price changes are expected to be greater in cities that are more regulated on land supply or have less developable land shares.

---

[8] $Newly\ added\ land_{03-12} = \log\left(\sum_{t=2003}^{2012} N_{i,t}\right)$, where $N_{i,t}$ represents an annual quota for the amount of newly- added construction- land that is converted from farmland and requisitioned land.

[9] Controls include: construction cost 2001, income per capita 2001, population 2001, pop density 2001, East dummy, January temperature. Housing price growth may be driven by increasing in construction cost. Income may influence demand in the market due to affordability. In addition, population and pop density affects the rate of house-hold formation, which in turn, is also one of key determinants for housing demand. East dummy variable is used to represent the eastern coastal cities which have had more economic growth and more stringent government regulation than other areas.

The estimation of quantity follows the same approach by using the logarithm of the change in the number of newly-built apartment units between 2005 and 2012 as the dependent variable. Conditional on demand shocks, long changes in quantity are expected to be smaller in the areas where housing supply is more responsive:

$$\Delta\log(\text{Newly\_built units})_{i,05-12} = \beta_0 + \beta_1 devshr_i + \beta_2 Newly\ increased\ land_i + \beta_3 X_i + \mathrm{u}_i. \quad (6)$$

Newly-built apartment units refer to the number of residential units for which construction was completed in the current year. Data on newly-built units are available starting in 2005, therefore the regression only covers the 2005-2012 period.

The allocation of newly-increased land is to some extent exogenous[10] to housing price, but it may be correlated with a city's size and unobservable demand shocks. To solve the potential endogeneity problem, I use two instrumental variables: Bartik's shift-share variable (Bartik, 1991) and farmland in the year 2001. The Bartik instrument is a measure of local labor demand using local industry employment shares in the beginning year weighted by national employment growth rate in each industry.[11] The Bartik instrument has been used to model exogenous sources of demand shocks in the urban economics literatures (Glaeser, Gyourko, and Sakes (2006) and Saiz (2010)). The second instrument for newly-increased land is farmland. One aim of the land quota system in China is to protect farmland from exploitation. The central government imposed a policy that requires the conversion of agriculture land to construction land to be offset by conversion of other land to agricultural use (Lichtenberg and Ding, 2011). I use the quantity of farmland (in hectares) in the year of 2001, two years before my sample period starts, to allow the measure to be exogenous. The first-stage joint test F-statistic is 10.38.

[10] The central government is in charge of approving the conversion of farmland to construction land and the supply of newly added construction land. Local governments are responsible for putting land on the market through open auction and are given the autonomy to decide the purpose the land is to be used for (manufacturing, commercial and residential) (Cai, 2011).

[11] I construct the Bartik instrument using employment shares at the 2 digit SIC-level in 2000 weighted by the national growth rate (2000-2010). Data sources: China population census 2000, 2010.

The specification could be pushed further to find an estimate of the price elasticity of housing supply, following a regression model developed by Saiz (2010). Appendix Table C displays the results. The associated supply elasticities are greater than one, implying an elastic response, ranging from 1.77 to 4.03. A negative coefficient value on the interaction term ($devshr \cdot \Delta \log Q_i$) implies that natural land availability has mediated the impact of the demand shock on the housing price, though the coefficient is insignificant.

2.6 Empirical results and analysis

This paper examines whether markets with large land availability (less geographical and/or man-made constraints) will experience small price gains and build more housing. Table 2 presents the results of estimating the regression models (5) and (6). The variables shown on the top row are dependent variables. Column 1 of table 2 displays the result of regressing each city's price changes between 2003 and 2012 on geographical constraints. A negative coefficient value for developable land share implies that prices move less in the places that have more naturally land available conditional on observable demand shocks. Column 2 incorporates all the variables specified in equations (5). The coefficient value of geographical constraints is in line with the column 1. A negative value of -0.382 implies that a one standard deviation change in the developable land share (which equals 17 percentage points) is associated with 6.74 % reduction in real price changes. The coefficients on newly increased land are statistically insignificant in both OLS (column 2) and 2SLS (column 3) estimations. The results imply that the supply condition proxy for government regulation on land supply is unresponsive to price changes. The land quota system is planned at the central level in China. The central government is in charge of approving the supply of newly added construction land. Therefore, land supply in China can be considered as quasi-exogenous. It is the demographic variable, income per capita, which accounts for a significant fraction of cross sectional variation among the demand controls. In principle, housing supply is a function of construction supply and land supply. The existing literature reports that appreciation of construction cost cannot account for the large increases in house prices in China (Wu, Gyourko, and Deng, 2015). The result in my paper is consistent with the finding in the existing literature that construction cost is not strongly correlated with price changes. Column 3 shows the results with instruments, which are in line with column (2).

Columns 4-6 reported in table 2 examine quantity, using the logarithm of the number of newly-built apartment units between 2005 and 2012 as the dependent variable. In column 4 the coefficient value of developable land share is positive and significant, suggesting that cities with more available land are likely to build more housing. The column 5 includes all the variables in equation (6). Among controls, population density and construction cost contribute most to differential changes in quantity supply. Column 6 displays the results with instruments. Both OLS and 2SLS estimations show that newly increased land does not affect quantity.

Table 2: Housing Price, Construction and Supply Conditions in 35 Chinese Cities

	ΔLog(Price) 03-12			ΔLog(Newly-built units) 05-12		
	OLS	OLS	2SLS	OLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Devshr	−0.332*	−0.382*	−0.376*	1.059*	0.839*	0.783
	(0.180)	(0.205)	(0.191)	(0.581)	(0.476)	(0.640)
Newly added $\text{land}_{03\text{-}12,\text{hectare}}$		0.031	0.007		−0.062	0.145
		(0.036)	(0.076)		(0.149)	(0.327)
Controls	No	Yes	Yes	No	Yes	Yes
R-squared	0.110	0.376	0.368	0.089	0.400	0.354
Observations	35	35	35	35	35	35

Notes: Columns (1)-(3) show results for change in price. Columns (4)-(6) show results for change in newly-built apartment units. $Newly\ increased\ land_{03-12} = \log \sum_{t=2003}^{2012} N_{i,t}$, where $N_{i,t}$ represents an annual quota for the amount of newly added construction land that is converted from farmland and requisitioned land. Column (3) and (6) show results for 2SLS models that allow for the potential endogeneity of newly added land. The instruments used for newly added land are a Bartik shift-share variable and farmland in 2001. Controls include: income per capita 2001, population 2001, popdensity 2001, East dummy, construction cost, January temperature. Table of summary statistics is in appendix. Heteroskedasticity-robust standard are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%.

2.7 Conclusion

In China, high housing prices have become a long run phenomenon and have soared in recent years. Many empirical studies examine various aspects of the Chinese housing market. My paper focuses on the role that housing supply has played in explaining the variation in housing prices and quantities, controlling for demand shocks. Using data from 35 Chinese cities, I examine how geographical and man-made constraints are related to the change in housing prices and quantities between 2003 and 2012. While the sample size is small, the results are suggestive about how the Chinese housing market has been working since 2003. The results imply that cities with less developable land have experienced greater price appreciation in the past decade and the quantity response is less in those places. My findings suggest geography matters in Chinese housing markets where land is discretely allocated by the government. In cities where there is more land naturally available, the government may be less concerned about the loss of farmland and be more permissive with development[12]. Moreover, my findings imply that the allocation of land use via government decision is quasi-exogenous to changes in housing price and quantity, suggesting decisions of governments about land supply may not dependent on housing prices. As data becomes available on more Chinese cities, the framework developed here can be used to enrich our understanding of Chinese housing markets.

[12] I thank the referee for providing this insight.

References

Bartik, T. J. (1991). Who Benefits from State and Local Economic Development Policies? *Journal of the American Planning Association*, *58*(2): 258-259.

Cai, M. (2011). Local Determinants of Economic Structure: Evidence from Land Quota Allocation in China. Working Paper.

Cao, G., Feng, C., and Tao, R. (2008). Local land finance in China's urban expansion: challenges and solutions. *China & World Economy,* 16(2):19-30.

Chan, W.K., (2013). The Problem with Chinas Urban Population Data, East Asia Center. University of Washington. Retrieved 28 April 2013.

Davido, T. (2013). Supply elasticity and the housing cycle of the 2000s. *Real Estate Economics*, 41(4):793-813.

Deng, Y., Gyourko, J., and Wu, J. (2012). Land and housing price measurement in China. *In Property Markets and Financial Stability, Proceedings of a Conference, Reserve Bank of Australia, Sydney* (pp. 13-43).

Ding, C. and Lichtenberg, E. (2011). Land and urban economic growth in China. *Journal of Regional Science,* 51(2):299-317.

Dong, Y. (2016). A note on geographical constraints and housing markets in China. *Journal of Housing Economics*, 33, 15-21.

Du, H., Ma, Y., and An, Y. (2011). The impact of land policy on the relation between housing and land prices: Evidence from China. *The Quarterly Review of Economics and Finance*, 51(1):19-27.

Fu, Y., Zheng, S., and Liu, H. (2009). Population growth across Chinese cities: Demand shocks, housing supply elasticity and supply shifts. *Housing Supply Elasticity and Supply Shifts* (May 1, 2008).

Fu, Y., Zheng, S., and Liu, H. (2012). Structural Estimates of Housing Supply Elasticity across Chinese Cities. Working Paper.

Glaeser, E. L., Gyourko, J., and Saiz, A. (2008). Housing supply and housing bubbles. *Journal of Urban Economics*, 64(2): 198-217.

Glaeser, E. L., Gyourko, J., and Saks, R. E. (2005). Why have housing prices gone up? *American Economic Review,* 95(2):329-333.

Gyourko, J. and Saiz, A. (2006). Construction costs and the supply of housing structure. *Journal of Regional Science,* 46(4):661-680.

Gyourko, J., Saiz, A., and Summers, A. (2008). A new measure of the local regulatory environment for housing markets: The wharton residential land use regulatory index. *Urban Studies*, 45(3):693-729.

Leman, E., (2005). Metropolitan Regions: New Challenges for an Urbanizing China. *World Bank/IPEA Urban Research Symposium* (Vol. 4).

Lichtenberg, E., & Ding, C. (2008). Assessing farmland protection policy in China. *Land Use Policy*, *25*(1), 59-68.

Lichtenberg, E. and Ding, C. (2009). Local officials as land developers: Urban spatial expansion in China. *Journal of Urban Economics*, 66(1):57-64.

Liu, T., & Huang, M. (2016). Land supply and money growth in China. *International Review of Economics & Finance*, 42:327-338

Mian, A. and Sufi, A. (2009). The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *The Quarterly Journal of Economics*, 124(4):1449-1496.

Peng, L. and Thibodeau, T. G. (2012). Government interference and the efficiency of the land market in China. *The Journal of Real Estate Finance and Economics*, 45(4):919-938.

Quigley, J. M. and Raphael, S. (2005). Regulation and the high cost of housing in California. *American Economic Review*, 95(2):323-328.

Saiz, A. (2008). On local housing supply elasticity. *Available at SSRN 1193422*.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics,* 125(3):1253-1296.

Tan, R., Qu, F., Heerink, N., and Mettepenningen, E. (2011). Rural to urban land conversion in China: how large is the over-conversion and what are its welfare implications? *China Economic Review,* 22(4):474- 484.

Wang, S., Chan, S. H., and Xu, B. (2012). The estimation and determinants of the price elasticity of housing supply: Evidence from China. *Journal of Real Estate Research*, 34(3):311-344.

Wang, Y., Tu, Y., & Li, W. (2014). Neighborhood land-use patterns and housing prices: evidence from the Beijing housing market. Working Paper.

Wheaton, W. C. (2004). Commuting, congestion, and employment dispersal in cities with mixed land use. *Journal of Urban Economics*, 55(3):417-438.

Wu, J., Gyourko, J., and Deng, Y. (2015). Evaluating the risk of Chinese housing markets: What we know and what we need to know. Working Paper.

Xiao, Y. and Zhao, J. (2015). Fixing China's distorted urban land quota system. *Paulson Policy Memorandum*. *Retrieved*

Zheng, S. and Kahn, M. E. (2008). Land and residential property markets in a booming economy: New evidence from Beijing. *Journal of Urban Economics*, 63(2):743- 757.

Appendix Table A: China Cities Developable Land Shares (%)

City	Slope>15^{o} area	Sea area	Internal water area	Developable land share
Shenyang	4.09	1.35	-	93.65
Zhengzhou	4.02	-	2.68	93.29
Tianjin	0.26	-	6.92	92.82
Harbin	4.91	-	2.38	92.70
Shijiazhuang	6.82	-	0.73	92.45
Hefei	1.21	-	6.74	92.05
Changchun	6.68	-	1.88	91.44
Changsha	9.48	-	2.09	88.43
Nanchang	5.51	1.47	5.26	87.76
Guangzhou	6.23	1.84	4.29	87.64
Yinchuan	9.95	-	2.53	87.52
Beijing	10.56	-	1.88	87.11
Chengdu	14.68	-	0.31	85.00
Xi’an	16.40	-	0.56	83.04
Shanghai	0.48	16.93	1.12	81.48
Wuhan	2.49	0.38	15.75	81.39
Jinan	16.94	0.51	1.37	81.18
Nanjing	5.69	9.20	4.42	80.69
Hangzhou	19.27	0.09	2.19	78.44
Urumqi	20.16	-	1.50	78.35
Huhehot	24.18	-	-	75.82
Nanning	25.44	0.08	1.22	73.26
Chongqing	29.46	-	2.96	67.58
Taiyuan	34.17	-	0.03	65.80
Guiyang	36.19	-	-	63.80
Haikou	0.18	41.03	-	58.78
Ningbo	26.06	16.47	1.27	56.19
Kunming	40.28	-	6.49	53.22
Xining	48.20	-	-	51.80
Fuzhou	44.11	2.58	4.05	49.27
Xiamen	13.89	36.17	0.95	48.97
Shenzhen	16.96	33.12	1.06	48.86
Qingdao	6.16	47.8	0.27	45.73
Lanzhou	53.58	-	0.73	45.69
Dalian	5.89	60.21	0.16	33.73

Notes: Developable land share = 100(1－(slope15/total area)－(sea area/total area)－(internal water area/total area)),developable land share represents the share of land that is suitable for housing construction within 35 kilometers from a city center, and is derived from ASTER Global Digital Elevation Model (30m resolution version 2, Oct. 2011) and the author’s own GIS calculations. Internal water shapefile provided by City University of New York, Internal ESRI Data.3

Appendix Table B: Summary Statistics for Key Variables for 35 Chinese Cities

Variable	Mean	Std. Dev.	Min	Max	Obs.
ΔLog(Price) (03-12)(%)	85.48	17.66	49.76	130.69	35
Developable land share	0.735	0.176	0.337	0.936	35
Newly added land(03-12)(hectare)	9020.23	5901.81	795.45	20382.66	35
Population (10 thousands)(2001)	629.38	523.38	57.34	3091.00	35
Real income per capita (2001)	46833.52	9583.139	34616.11	76848.05	35
Population density(2001)	620.08	469.71	121	2430	35
Construction Cost (2001)	1353.29	617.69	810.62	4070.40	35
Bartik shift-share (%)	6.55	6.06	− 4.45	19.25	35
January Temperature (Centigrade)	5.78	8.28	−12.65	20.92	35
Farmland(2001)(thousands hectare)	345.29	351.35	3	1583	35

Data sources: China Real Estate Statistics Yearbook (2000-2013), China Cities Statistics Yearbook (2000-2013), China Population Census 2000, 2010. Construction cost is measured in yuan/square meters, real terms

Appendix Table C: Inverse Housing Supply with Natural Constraints (35 Chinese Cities)

	ΔLog(P): 2000-2010			
	OLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)
ΔLog(Q)$_{\text{household}}$	0.248 **	0.681*	0.584**	0.682
	(0.126)	(0.363)	(0.273)	(0.533)
Developable land·ΔLog(Q)$_{\text{household}}$		−0.764		−0.235
		(0.577)		(0.854)
East	Yes	Yes	Yes	Yes
R-squared	0.167	0.197	0.088	0.119
Observations	35	35	35	35

Notes: The estimating equation is $\Delta \log P_i = \beta^s \cdot \Delta \log Q_i + \beta^{Geography} \cdot devshr \cdot \Delta \log Q_i + R_i + u_i$. The independent variable $\Delta \log Q_i$ is the change in housing demand (the log of the number of family households), between 2000 and 2010, the data on family households (unit: household) is obtained from Census data 2000 and 2010. Time period is determined by availability of 100%-count census data. ΔLog(P) is measured between 2000-2010 with respect to household. The instruments used for demand shocks (Δlog(Q)$_{\text{household}}$) are Bartik shift-share and January temperature. The first-stage joint test F-statistic is 10.37. The implied supply elasticity is 4.032 (1/0.248) in column 1 and 1.773 (1/0.584) in column 3. A negative coefficient value for of the interaction term ($devshr \cdot \Delta \log Q_i$) implies that natural land availability has mediated the impact of the demand shock on the housing price. More land availability shifts down the supply curve. R_i: region fixed effect. *significant at 10%, **significant at 5%, ***significant at 1%.