

NBER WORKING PAPER SERIES

ENTREPRENEURSHIP AND URBAN GROWTH:  
AN EMPIRICAL ASSESSMENT WITH HISTORICAL MINES

Edward L. Glaeser  
Sari Pekkala Kerr  
William R. Kerr

Working Paper 18333  
<http://www.nber.org/papers/w18333>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
August 2012

Comments are appreciated and can be sent to [eglaeser@harvard.edu](mailto:eglaeser@harvard.edu), [skerr3@wellesley.edu](mailto:skerr3@wellesley.edu), and [wkerr@hbs.edu](mailto:wkerr@hbs.edu). Glaeser and W. Kerr are affiliates of the NBER. We thank Ajay Agrawal, Hoyt Bleakley, Mercedes Delgado, Xavier Giroud, Rick Hornbeck, Larry Katz, James Lee, John McHale, Debarshi Nandy, Tom Nicholas, László Sándor, Curtis Simon, Will Strange, Adam Storeyguard, Bob Strom, and seminar participants for very helpful comments; Alex Field, Alex Klein, and Gavin Wright for their guidance with respect to historical data sources; Chris Hansen for assistance with the IVQR methodology; the Sloan Foundation and The Taubman Center for State and Local Government for financial support; and Kristina Tobio for excellent research assistance. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Support for this research at the Boston RDC from NSF (ITR-0427889) is also gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2012 by Edward L. Glaeser, Sari Pekkala Kerr, and William R. Kerr. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines

Edward L. Glaeser, Sari Pekkala Kerr, and William R. Kerr

NBER Working Paper No. 18333

August 2012

JEL No. L0,L1,L2,L6,N5,N9,O1,O4,R0,R1

**ABSTRACT**

Measures of entrepreneurship, such as average establishment size and the prevalence of start-ups, correlate strongly with employment growth across and within metropolitan areas, but the endogeneity of these measures bedevils interpretation. Chinitz (1961) hypothesized that coal mines near Pittsburgh led that city to specialization in industries, like steel, with significant scale economies and that those big firms led to a dearth of entrepreneurial human capital across several generations. We test this idea by looking at the spatial location of past mines across the United States: proximity to historical mining deposits is associated with bigger firms and fewer start-ups in the middle of the 20th century. We use mines as an instrument for our entrepreneurship measures and find a persistent link between entrepreneurship and city employment growth; this connection works primarily through lower employment growth of start-ups in cities that are closer to mines. These effects hold in cold and warm regions alike and in industries that are not directly related to mining, such as trade, finance and services. We use quantile instrumental variable regression techniques and identify mostly homogeneous effects throughout the conditional city growth distribution.

Edward L. Glaeser  
Department of Economics  
315A Littauer Center  
Harvard University  
Cambridge, MA 02138  
and NBER  
eglaeser@harvard.edu

William R. Kerr  
Harvard Business School  
Rock Center 212  
Soldiers Field  
Boston, MA 02163  
and NBER  
wkerr@hbs.edu

Sari Pekkala Kerr  
Wellesley College  
106 Central Street  
Wellesley, MA 02481  
skerr3@wellesley.edu

# 1 Introduction

Economists and policy makers often argue that urban success depends upon a city's level of entrepreneurship. This claim was famously made in Chinitz's (1961) comparison of New York and Pittsburgh, and it is more recently invoked by Saxenian (1994) when contrasting the regional performance of Boston and Silicon Valley. More systematic empirical evidence confirms these famous case studies. For example, Glaeser et al. (1992) find a strong correlation between small establishment size and subsequent employment growth across sectors within U.S. cities. Glaeser, Kerr and Ponzetto (2010) also document the strength of this relationship when modelling entrepreneurship through start-up employment shares. Similar conclusions are reached recently by Delgado, Porter and Stern (2010a,b), Rosenthal and Strange (2003, 2010), and Gennaioli et al. (2012).

Figure 1 provides representative graphs from this work. These patterns are frequently taken as evidence that entrepreneurship is an important ingredient for local job growth. While the empirical association is quite visible, there are clearly many factors that jointly influence initial entrepreneurship levels and subsequent growth of cities (e.g., regional growth trends, local public policies). Without identifying exogenous sources of variation for entrepreneurship, it is premature to make strong claims that entrepreneurship causes urban growth.<sup>1</sup>

We tackle this problem by using an idea suggested in Chinitz's original account. Chinitz claimed that Pittsburgh's dearth of entrepreneurs in the 1950s reflected its historical concentration in steel, which in turn reflected proximity to large deposits of coal and iron ore (White, 1928). The steel industry has significant returns to scale, and Chinitz argued that its presence crowded out more entrepreneurial activities. This left Pittsburgh with an abundance of company men but few entrepreneurs. Moreover, Chinitz emphasized how this dampening of entrepreneurship comes through both static factors (e.g., access to inputs for new businesses) and dynamic factors (e.g., the transmission of skills and attitudes from parents to children). Chinitz's hypothesis was, in a sense, an early conception of the natural resource curse, which in this case, operates through large, resource-intensive activities crowding out the entrepreneurial activity which generates long-term growth.<sup>2</sup>

We systematically investigate the connection between historical mineral and coal deposits and modern entrepreneurship. There are returns to scale in many extractive industries and their industrial customers, not just coal and steel. The process of bringing ores out of the earth is a capital-intensive operation that often benefits from large-scale operations. Transforming and transporting ores also typically requires large machines and production facilities. Therefore,

---

<sup>1</sup>Further progress has been made in establishing causal links of entrepreneurial finance to industry or city growth (e.g., Kortum and Lerner, 2000; Samila and Sorenson, 2011). It is striking, however, that similar progress has not been achieved for entrepreneurship overall. Agrawal et al. (2012) explore firm size distribution and the nature of innovation in cities, emphasizing a dual role of a large incumbent firm and many start-ups.

<sup>2</sup>If entrepreneurs generate positive externalities relative to employees of U.S. Steel, then this crowding out can be socially harmful, but the Chinitz hypothesis can still operate without any market failures. van der Ploeg (2011) surveys work in a related literature on natural resources and country success.

we hypothesize that cities with a historical abundance of nearby mineral and coal mines will have developed industrial structures with systematically larger establishments and less entrepreneurship. These early industrial traits can in turn influence modern entrepreneurship through persistence and intergenerational transmissions that we elaborate on further below.

We use the existence of mineral and coal deposits in 1900 as our measure of the returns to mining around a city. These data come from the historical records of the U.S. Geological Survey and economic censuses at the time. Figure 2 is a representative map. We demonstrate that a city's historical proximity to mineral and coal deposits is strongly correlated with larger average establishment size for manufacturing in 1963 and subsequently. These deposits are also associated with larger average establishment size in quite unrelated industries in the 1970s and 1980s (initial years for sectors are determined by our Census Bureau data). While the relationship is most pronounced in industries that have more occupational overlap with mining, historical deposits are associated with larger establishment sizes throughout the city. These patterns are very similar for other measures of modern entrepreneurship like local employment in start-up firms.

With this background, we use historical mineral and coal deposits as instruments for modern entrepreneurship. We continue to find a strong connection between a city's initial entrepreneurship and subsequent economic growth. A one standard deviation decrease in average initial establishment size for a city is associated with a 0.61-0.88 standard deviation higher employment growth between 1982 and 2002. Similarly, a one standard deviation increase in the initial employment share of start-up firms is associated with a 0.25-0.35 standard deviation increase in urban employment growth over the next two decades. These ranges reflect variations in distance bands around cities for the impact of mines. The instrumented elasticities for average initial establishment size are similar to ordinary least squares estimates; the instrumented elasticities for initial employment share of start-up firms are larger than ordinary least squares estimates.

Our primary concern with these results is that mineral and coal deposits are likely associated with other variables that can impact economic growth. Unionization is a prime candidate (Holmes, 2006), but we can explicitly control for this variable. These correlations may also reflect a general decline in U.S. employment in extractive industries or the decline of Rust Belt regions. We address the first concern by separately considering industries that are quite different from mining, such as trade, services, and finance. We find that our results are, if anything, stronger for these sectors of the economy. Proximity to mines in 1900 predicts larger establishments, less entry, and less urban growth in trade, services, and finance today.

Sector decompositions do not address the possibility that our results simply reflect the general decline of cities that were initially built around natural resources. The decline of the steel industry in Pittsburgh did not just impact steel production, but also the financial and service firms that catered to that industry and its employees. We have two complementary approaches to test this concern. Our first approach is to focus on the United States' growing regions.

Manufacturing does not predict strong urban decline in the warmer regions of the United States, which have witnessed the most substantial urban growth over the past several decades, and yet we still find that historical mines predict dampened employment growth. Service industries that are highly agglomerated in a small number of areas are typically believed to be oriented towards national and international sales, rather than the local market. We also continue to find the negative connection between mines and employment growth effects in highly agglomerated industries that should be less dependent on local demand. These patterns continue to hold as well in warmer areas, although some sensitivity to the spatial range of the instruments is evident. We also show that our results are robust to including controls for the projected forward employment growth of the city based upon its initial industry composition and national growth trends for industries and the observed change in manufacturing employment for the city from 1963 to 1981.

Our second approach is more technical in nature but less dichotomous than grouping cities and industries. We implement the instrumental variable quantile regression method (IVQR) of Chernozhukov and Hansen (2004a, 2005, 2006). This econometric technique effectively estimates the instrumental variable regressions at various points throughout the city growth distribution, where growth is conditional on specified covariates such as climate, initial housing prices, regional fixed effects, and similar. We show that the impact of initial establishment size on subsequent employment growth is reasonably homogeneous throughout the conditional distribution. That is, entrepreneurship is linked to stronger subsequent employment growth in cities that are growing faster as well as those growing slower than what their initial traits would have predicted. To the extent that it differs by city growth, the connection of entrepreneurship to city growth is most important among cities that are underperforming in their growth.

In the last part of the paper, we consider several variations on our city growth measures that take advantage of the micro-data. We show similar results when measuring employment in 2002 contained in establishments that did not exist in 1982, finding stronger elasticities than our overall measures. We also quantify how higher initial entrepreneurship is linked to greater employment shares for entrants since 1982 throughout the establishment size distribution, with new employment being retained relatively more in larger establishments. Higher initial entrepreneurship in 1982 is also associated with lower average establishment ages in 2002 for the city, both generally and among the top 25 employers for the city. As we discuss later, these extensions and others suggest that the up-or-out process outlined by Haltiwanger et al. (2012) at the firm level when linking young firms to employment growth is also holding more systematically at the city level for urban growth dynamics.

These results and their stability suggest that mines influenced modern entrepreneurship with a much deeper foundation than U.S. regional evolution. Nevertheless, historical mineral and coal deposits are an imperfect instrument. They will have some correlation with other local variables besides entrepreneurship. Thus, our conclusions must be tentative. Yet empirical work on en-

trepreneurship and urban economics must begin identifying and exploiting exogenous sources of local entrepreneurship. Historical mines are one such instrument, imperfect as they may be. Our work represents a step towards identifying exogenous sources of variation in local entrepreneurship and using that variation to examine whether the strong correlations between city employment growth and entrepreneurship hold when removing the most worrisome endogeneity. The general conclusion from this exercise is that entrepreneurship is systematically related to local employment growth over the past three decades.

The plan of this paper is as follows. Section 2 reviews the Chinitz hypothesis and related literature. Section 3 introduces our data, and Section 4 establishes basic facts about entrepreneurship, average establishment size, and urban employment growth. Section 5 then describes in greater detail the first stage relationships between historical mineral and coal deposits and modern entrepreneurship. Section 6 presents the core instrumental variable results. Section 7 provides the extended employment growth results, and Section 8 concludes.

## 2 Entrepreneurship, Establishment Size and Mining

The core hypothesis of the literature on entrepreneurship and city growth is that some places are endowed with a greater number of entrepreneurs than others and that this endowment of entrepreneurial human capital influences economic success. Chinitz (1961) first formulated this hypothesis in his attempt to explain why post-war New York was experiencing more economic success than post-war Pittsburgh. In a sense, this entrepreneurship hypothesis is a close cousin of the literature relating local human capital levels to area development and growth.<sup>3</sup> While the latter human capital literature typically focuses on formal education as the measure of human capital, entrepreneurial skill is another important form of human knowledge that seems a priori as likely to explain area success as any other type of skill.

The literature on entrepreneurship and local economic growth typically uses two different measures of entrepreneurship, neither of which are perfect. Perhaps the most common choice is average establishment size, which is readily available in public data sources like County Business Patterns. Small establishments would seem to be a natural measure of the ratio of the number of establishment heads, who may be entrepreneurs, to employees. Micro-data studies, on the other hand, often emphasize that young and entering establishments generate more job growth than small establishments.<sup>4</sup> Thus, a second measure of entrepreneurship is the share of local employment that is in new start-up firms. While the latter metric captures more of the dynamic nature of entrepreneurship, it also frequently requires access to confidential micro-data. Nevertheless, these two measures are highly correlated with each other across cities, and both have

---

<sup>3</sup>For example, Glaeser et al. (1995), Simon (1998), Simon and Nardinelli (2002), and Gennaioli et al. (2012).

<sup>4</sup>For example, Davis, Haltiwanger and Schuh (1996), Haltiwanger, Jarmin and Miranda (2012), and Hurst and Pugsley (2012).

been shown to be correlated with local employment growth.<sup>5</sup>

Glaeser et al. (1992) find a link between small establishment size and sectoral employment growth between 1956 and 1988. Their basic approach is to look at city-industries—industrial groups within cities—and they observe that city-industries with smaller average establishment sizes grew more rapidly. Glaeser, Kerr and Ponzetto (2010) follow this work using the Longitudinal Business Database and find that the correlation is extremely strong and robust. The patterns hold with city and industry fixed effects and across a broad range of industries and regions. They also observe that areas with small establishment sizes do not seem to have higher returns to entrepreneurship, which supports the idea that cities differ sharply in their supply of entrepreneurs.<sup>6</sup>

But while it is clear that some cities and city-industries have much larger average establishment sizes, and that employment growth is lower where establishments are bigger, it is less clear why establishment sizes differ spatially. Glaeser, Kerr and Ponzetto (2010) interpret their results as meaning that clusters of entrepreneurship exist, but they are unable to explain why they exist where they do. Without adequate sources of exogenous variation in entrepreneurship, it is impossible to be sure that the measured growth effects of entrepreneurship really represent the causal effect of entrepreneurship or whether there are other factors that lead cities to have both more growth and more entrepreneurship.

Our approach to this problem starts with Chinitz’s claim that industrial history drives the level of entrepreneurship in a city. Chinitz argues that New York’s historical garment industry—the nation’s largest post-war industrial cluster—was a natural training ground for entrepreneurs. The garment trade had few serious fixed costs or scale economies, and as a result there were a large number of small entrepreneurs in the industry. Chinitz argued that this entrepreneurship in turn influenced neighboring industries.

Indeed, there are many anecdotes about entrepreneurs who began in the garment industry and then branched into other industries (or bred entrepreneurial children). For example, A. E. Lefcourt was New York’s greatest skyscraper builder in the years before the Great Depression. Lefcourt got his start in the garment trade, where he was able to scrape together enough capital from his savings and by borrowing from his customers to buy a garment company from his boss at the age of 25. The father of Sanford Weill, an entrepreneurial engine in New York’s finance industry from the 1960s to the 1990s, also started as a garment entrepreneur. These stories support Chinitz’s contention that entrepreneurial human capital may actually be transmitted from parent to child.

By contrast, Chinitz depicts Pittsburgh as a city of company executives who did not want

---

<sup>5</sup>Self-employment is a third possible measure of entrepreneurship. While it is correlated with average establishment size across metropolitan areas (Glaeser and Kerr, 2009), it is considered to be a very noisy measure and has little correlation with economic growth. As such, we do not use it in this study.

<sup>6</sup>Acs and Armington (2006) provide a broad overview of U.S. spatial patterns for entrepreneurship and economic growth. Ghani, Kerr and O’Connell (2010) document similar patterns across regions and industries in India. Miracky (1993) further extends the work of Glaeser et al. (1992).

nor could have inculcated entrepreneurial talents in their children. Chinitz suggests the roots of this big company mentality came from Pittsburgh's dominant steel industry. The steel industry was dominated by a few large firms, most notably U.S. Steel, which accounted for 66 percent of ingot production in 1901 and 42 percent in 1925 (Stigler, 1925).<sup>7</sup> U.S. Steel, of course, had its roots in the scrappy start-ups of Andrew Carnegie and others, but by the early decades of the 20th century, it had become synonymous with corporate bigness. Chinitz (1961) then argues:

My feeling is that you do not breed as many entrepreneurs per capita in families allied with steel as you do in families allied with apparel, using these two industries for illustrative purposes only. The son of a salaried executive is less likely to be sensitive to opportunities wholly unrelated to his father's field than the son of an independent entrepreneur. True, the entrepreneur's son is more likely to think of taking over his father's business. My guess is, however, that the tradition of risk-bearing is, on the whole, a more potent influence in broadening one's perspective.

In Chinitz's view, the "salaried executives" of U.S. Steel were just less likely to inculcate entrepreneurial talents and inclinations in their children, which in turn made Pittsburgh less entrepreneurial for years to come.

Chinitz certainly seems to be right about intergenerational transmission of entrepreneurship (Blau and Duncan, 1967; Niittykangas and Tervo, 2005). Hout and Rosen (2000) document that "the primary family factor affecting an individual's self-employment status is the self-employment status of his or her father." They show that self-employment rate for sons of self-employed fathers is about twice as high as the self-employment rate for sons of employees. The intergenerational transmission of entrepreneurial human capital makes it possible that industrial history could still impact the level of entrepreneurship today. The likelihood of this persistence is supported by empirical studies that show that entrepreneurs are more likely to be from their region of birth than wage workers, and that local entrepreneurs operate stronger businesses.<sup>8</sup>

Chinitz further documents a number of reasons why the broader ecosystem of entrepreneurship can be depressed by large incumbent firms. In addition to the intergenerational mechanism, Chinitz discusses social standing more broadly, suggesting that an "aura of second-class citizenship" surrounds entrepreneurship in cities dominated by big firms. Chinitz also notes capital constraints: small firms are more likely to redeploy capital in their local area than large firms, and financial institutions are also more likely to serve small firms in cities with more small firms. These patterns have been subsequently observed in multiple entrepreneurial finance studies. Chinitz further emphasizes labor constraints, as large firms are more likely to locate out of the center city, which makes spousal employment harder. Finally, and perhaps most famously,

---

<sup>7</sup>Stigler's famous piece on U.S. Steel emphasizes that the creation of this company brought massive returns to investors because of its ability to exploit monopoly power.

<sup>8</sup>For example, Figueiredo, Guimaraes and Woodward (2002), Michelacci and Silva (2007), and Dahl and Sorenson (2007). See also Whyte (1956).

Chinitz emphasizes access to intermediate goods. Small firms have many needs that must be satisfied by the local economy. Large incumbent firms often source inputs internally or at a distance. This can depress external supplier development. Moreover, similar to capital providers, it then becomes harder for new entrants to gain the attention of existing suppliers that are serving large firms in the area. These additional factors also make it harder for entrepreneurship to get underway in a city with large incumbent firms.

To find exogenous variation in a city's industrial past, we turn to mineral and coal mines. The U.S. Geological Survey has been documenting the existence of such deposits for over a century, and we are able to determine whether deposits exist near any given city. We hypothesize that these deposits were generally associated with bigger establishments and firms, just as coal mines were with U.S. Steel in Pittsburgh, and that those bigger establishments crowded out smaller enterprises and entrepreneurship.

Why would mines generally be associated with larger establishments? Mining itself appears to have substantial returns to scale, probably because of the large fixed investments required to drill, mine and ship heavy products like ore and coal. In 2008, County Business Patterns documents that the average establishment size across the entire United States is fewer than 16 people. By contrast, the average coal mining establishment has 74 people. The average iron ore mining establishment has 209 workers, and the average establishment in copper, nickel, lead, and zinc mining has 193 workers. It certainly appears that mining itself is conducive to large establishments, perhaps even more so than the documented accounts for coal mining.<sup>9</sup>

Pittsburgh's example suggests that manufacturing establishments that then use the products of mines are also large, perhaps because industries that use large amounts of coal or ores have large scale economies associated with big plants. In 2008, the average establishment in primary metal manufacturing had 85 employees, which is more than double the 40 employee national average for manufacturing as a whole. As such, it is plausible that an abundance of mineral and coal deposits led to large establishments in a particular area and that these large establishments meant that typical workers became skilled at working in big firms, not at starting their own companies.<sup>10</sup>

Our identification strategy builds on the exogenous spatial distribution of mineral and coal deposits in 1900. We first link these deposits to average establishment sizes and entrepreneurship in the 1960s and onwards. If Chinitz is right that big firms reduce the stock of entrepreneurial capital, then these deposits should lead to larger average establishment sizes in closely related industries, such as primary metal manufacturing, and also in less-related sectors like services and finance. We then investigate whether the places and sectors that have large average establishment

---

<sup>9</sup>In 1919, the average employee counts are similarly high: all mines (77), anthracite coal mines (508), bituminous coal mines (82), and iron ore mines (158). Calculations are made using the 1930 Statistical Abstract of the United States, Table 733.

<sup>10</sup>Related evidence on spin-outs includes Elfenbein et al. (2010), Gompers et al. (2005), and Klepper and Sleeper (2005).

sizes—because of proximity to mineral and coal deposits—experience less growth during the modern era.

Along these lines, it is important to note that proximity to historical mines provided past benefits to cities. Indeed, cities may have been founded precisely to exploit these deposits. As a simple calculation, regressions of log average household income and log city population from the 1950 Census of Population on coal production per capita in 1901 within 500 miles (Day, 1901; Haines, 2005) yields coefficients of 0.047 (0.010) and 0.335 (0.063), respectively, when controlling for regional fixed effects. These positive elasticities have since dissipated, to where a similar exercise using the 2000 Census of Population yields coefficients of 0.005 (0.009) and -0.009 (0.140), respectively. We seek to identify the extent to which this historical legacy from mining influenced local rates of entrepreneurship that appear very important for recent urban growth.

### 3 Data Description and Empirical Approach

This section describes our core data sources and empirical methodology. We develop our urban growth and entrepreneurship metrics through confidential data housed by the US Census Bureau. Our primary data source is the Longitudinal Business Database (LBD). The LBD provides annual observations for every private-sector establishment with payroll from 1976 onward. The only excluded sector is agriculture, forestry and fishing. In addition, we draw some statistics from the Census of Manufacturers, which extends back to 1963. Unfortunately, data for other sectors are only available starting in 1976.

The Census Bureau data are an unparalleled laboratory for studying the industrial structure of U.S. firms. Sourced from U.S. tax records and Census Bureau surveys, the micro-records document the universe of establishments and firms rather than a stratified random sample or published aggregate tabulations. In addition, the LBD lists physical locations of establishments rather than locations of incorporation, circumventing issues related to higher legal incorporations in states like Delaware.

The comprehensive nature of the LBD also facilitates complete characterizations of entrepreneurial activity by cities, industries, types of firms, and so on. Each establishment is given a unique, time-invariant identifier that can be longitudinally tracked. This allows us to identify the year of entry for new start-ups or the opening of new plants by existing firms. We define entry as the first year in which an establishment has positive employment. Second, the LBD assigns a firm identifier to each establishment that facilitates a linkage to other establishments in the LBD. This firm hierarchy allows us to separate new start-ups from facility expansions by existing multi-unit firms.

During a representative year, 1997, the data include 108 million workers and 5.8 million establishments. During the 1990s, there were on average over 700,000 entering establishments

each year that jointly employed more than seven million workers. The average start-up included ten workers, and notably there were very few entering mining establishments during this period (less than 0.5% of entrants).

Our core estimation examines urban growth and entrepreneurship from 1982-2002. We have manufacturing data going back to 1963, but we focus primarily on the period for which our data covers all sectors of the U.S. economy.<sup>11</sup> This will enable us to run regressions of the form

$$\ln \left( \frac{\text{Employment}_{c,2002}}{\text{Employment}_{c,1982}} \right) = \beta \cdot \ln(\text{Entrepreneurship}_{c,1982}) + \text{Other Controls}_c + \varepsilon_c, \quad (1)$$

where  $c$  indexes cities. We will use this same empirical design with industrial subsets of metropolitan areas. Our controls are taken from the urban growth literature and include initial employment, census division controls, and city-level variables like average January temperature, the share of adults with college degrees, initial housing prices, and similar.

The  $\beta$  coefficient describes the correlation of initial entrepreneurship and subsequent employment growth for the city. As in much of the previous research in this area, we focus on growth of employment rather than growth in wages, since wage differences across areas should be limited by the mobility of workers across space.<sup>12</sup> Entrepreneurs may, in addition, be able to succeed by limiting the wages received by the workers, so per capita wage growth is not necessarily a sign of local entrepreneurial success.

Our core measures of entrepreneurship are average establishment size in 1982 and the share of employment in start-ups in 1982-1986. We take the average over several years for the second metric to smooth out business cycles and the data collection patterns of the Census Bureau, but this is not an important factor. Average establishment size is defined as the number of employees divided by the number of establishments. It includes both single-unit firms and multi-unit establishments. We define the share of employment in start-ups on an annual basis using the entry rate of new single-unit firms. This approach quantifies gross entry levels, rather than the net entry that would be observed by looking at changes in establishments between two points.

Table 1a provides summary statistics for cities and entrepreneurship related to our sample. Throughout this paper, we conduct our analysis at the metropolitan area level, but we use the convention of referring to metropolitan areas as cities to ease exposition. We likewise refer to industries within metropolitan areas as city-industries.<sup>13</sup>

---

<sup>11</sup>We start our estimations in 1982, rather than in 1976, to be conservative. The period before 1982 includes a substantial amount of economic change and restructuring. Including this period leads to stronger results than those we present below, but we want to be conservative in our approach. Also, the LBD currently extends to 2007. We find very similar results when looking at total city employment growth until 2007. The Census Bureau, however, moves from the SIC industry classification system to the NAICS system in 2002. As this transition complicates many of our sector-level decompositions, we end the sample period in 2002.

<sup>12</sup>Standard models that assume a spatial equilibrium predict that increases in productivity increase employment. Wages rise with either increases in productivity or with decreases in local amenities, but the connection between productivity and wage changes depends on the elasticity of housing supply. Moreover, if declining industries fire their younger, lower-wage workers first, we can see rising average wages in declining sectors.

<sup>13</sup>We define cities by mapping counties in the LBD to Primary Metropolitan Statistical Areas (PMSAs). We

The average city had about 230 thousand employees in 1982 among sectors covered by the LBD. We will generally consider two large subsectors of the economy: "mining, construction and manufacturing" (which should be directly influenced by mining opportunities) and "trade, finance and services" (which should not make any direct use of coal or mineral ores). On average, a little less than three-quarters of city employments are in trade, finance and services. The average city experiences employment growth of 0.36 log points, or 44 percent, from 1982-2002. Reflecting national industrial trends, this employment growth is much higher in trade, finance and services (0.49) than in mining, construction and manufacturing (0.06). The average establishment has 19 employees, with substantially larger establishment sizes in mining, construction and manufacturing (34) than in trade, finance and services (16). About three percent of employees in a city are in entering firms over the 1982-1986 period.

Table 1b shows the correlation between these different measures of entrepreneurship. The first column shows the correlation between average establishment size and other measures of entrepreneurship. The first two rows show the connection between overall establishment size and establishment size within the two subsectors. The correlation between the overall measure and the first ore-oriented subsector variable is 0.62; the correlation with average establishment size in trade, finance and services is 0.74. The second column shows that the correlation in average establishment size between the two subsector-level variables is more modest at 0.14 (although statistically significant at a 10% level).

The third row in the first column shows the robust correlation between our two measures of entrepreneurship. Average establishment size in a city has a -0.49 correlation with the city's share of employment in start-up ventures. That is, cities with smaller establishments also have more employment in entering establishments. The fourth column shows the relationship holds when instead counting the share of establishments in a city that are start-ups. The final row shows that we find almost identical results to average establishment size when instead looking at the employment share in establishments with fewer than 20 employees, which is to be expected. The strong correlation between start-up employment and average establishment size is the topic of Glaeser, Kerr and Ponzetto (2010), who take it to suggest the existence of clusters of entrepreneurship.

The next two columns show the relationship between average employment size in the two industrial groups and other measures of initial entrepreneurship. Average establishment size in mining, construction and manufacturing is robustly correlated with start-up shares in the other variables. The correlation between average establishment size in trade, finance and services and the start-up shares is much weaker. Our empirical results focus on average establishment

---

exclude cities in Alaska and Hawaii due to our spatial instrument variable estimations. We also exclude some small PMSAs that are not separately identified in the Census of Population (required for explanatory variables). Results below are robust to instead considering Consolidated MSAs. CMSAs are subdivided into PMSAs for very large metropolitan areas (e.g., Chicago has six PMSAs within its CMSA). A PMSA is defined as a large urbanized county or a cluster of counties that demonstrate strong internal economic and social links in addition to close ties with the central core of the larger area.

size and employment shares in start-up firms. We find very similar results when using these additional variants.

## 4 OLS Relationship of Entrepreneurship and Local Growth

### 4.1 City Growth Regressions

We quantify the basic relationship between local entrepreneurship and subsequent urban employment growth. Equation (1) is our core empirical specification, but we also report results for growth in total payroll and wages. Panel A in Table 2 shows results using average establishment size in 1982 as our measure of entrepreneurship, while Panel B uses the initial share of employment in start-ups. Estimations are unweighted, have 291 observations, and report bootstrapped standard errors. We find smaller standard errors when clustering by the nine census divisions. To guard against excessive outliers, we winsorize variables at their 2% and 98% values.

The first regression in Panel A shows the strong negative relationship between employment growth over 1982-2002 at the metropolitan area level and initial establishment size. A one standard-deviation increase in 1982 establishment size is associated with a 0.57 standard-deviation decrease in the growth of employment over the ensuing 20 years. Panel B finds that one standard-deviation increase in the share of initial employment in start-ups is associated with a 0.2 standard-deviation increase in urban employment growth over the next 20 years. These effects are economically large and statistically significant, which is why it makes sense to further refine and test these correlations between entrepreneurship and local job growth.<sup>14</sup>

The second column shows that these coefficient estimates are essentially unchanged by including controls for the log level of initial employment in the city, its square, and fixed effects for the nine census divisions. This stability suggests that the correlations are not simply a product of mean reversion or differences in U.S. regional growth.

The third column shows that these coefficients are also robust to including standard controls for city growth from the urban growth literature: mean January and July temperatures, the 1970 share of workers with college degrees, the 1970 population level and density of the city, and 1970 housing prices. These factors control for documented phenomena like population growth over the last three decades in warm places and the rise of the skilled city. The fact that these controls have so little impact on our entrepreneurship measures suggests that these measures are unlikely to be proxying for core attributes of the urban area.<sup>15</sup>

Columns 4-6 repeat these results using payroll growth as the dependent variable. Some of the

---

<sup>14</sup>These results are quite robust to how the growth metric is defined, such as measuring growth relative to average city employment over 1982-2002 (e.g., Davis, Haltiwanger and Schuh, 1996). Similarly, non-parametric approaches that include indicator variables for quintiles of average establishment size demonstrate regular effects with the most substantial change occurring between the second and third quintiles.

<sup>15</sup>The results are further robust to additional covariates like Saiz's (2010) geographic features of cities or using hedonic regressions to model climate amenities. We lose several cities in these extension due to data availability, however, so we focus on the narrower set of controls.

coefficients are slightly smaller, but the overall picture remains the same. Metropolitan areas with more initial employment in start-ups or smaller average establishment size experienced faster payroll growth between 1982 and 2002. Other local controls have little effect on the core results.

In line with the symmetry of employment and payroll growth, Columns 7-9 confirm that initial entrepreneurship is not associated with subsequent wage growth nor declines. Entrepreneurship generates more job growth for cities, but not faster earnings growth for those employed. One interpretation of these results is that a spatial equilibrium exists across cities, and this equilibrium limits the tendency of any city's wages to rise much faster than its peers (Glaeser and Gottlieb, 2009). A second interpretation is that entrepreneurs have very lean operations that minimize labor costs, putting downward pressure on wage growth for workers. This latter effect could be due, for example, to entrepreneurs operating in more competitive environments.

## 4.2 Sample Decompositions

Table 3 examines patterns of employment growth within various subsets of our data, and Figures 3 and 4 provide a graphical analysis. The first column of Panel A repeats the total employment growth finding for initial average establishment size. Panel B then allows the treatment effect to differ by two broad regions of the United States. We group cities into cold cities, defined by having a mean January temperature less than 34 degrees, and warm cities. This cut-off point is approximately the median January temperature in the sample. Colder cities have a longer industrial history, experienced slower growth (or in some cases decline) over our time period, and include the complete Rust Belt. Entrepreneurship has a stronger association with city growth in colder regions of the United States. While the difference is statistically significant, its economic magnitude is small relative to overall effect. Panels C and D show a similar pattern for start-up employment shares.

Column 2-7 of Table 3 then repeat these specifications using various outcome variables. We define entrepreneurship at the city level, and we consider the types of industries in the cities where the employment growth is occurring. Column 2 examines employment growth in mining, construction and manufacturing. The results for average establishment size remain strong; the results for start-up employment shares become smaller and statistically insignificant. Column 3 shows that both measures are significant for trade, finance and services, although the start-up employment share has again lost some of its economic magnitude. At the city level, average establishment size appears the more robust correlate of subsequent employment growth across sectors.

Columns 4-6 separate employment growth by the degree of industrial agglomeration. We split industries by their national level of agglomeration as measured by the Ellison and Glaeser (1997) index. That index looks at the lumpiness of employment across space, correcting for the overall spatial distribution of economic activity and the tendency of industries with big

establishments to be more highly concentrated geographically. Our results are strongest for the most agglomerated industries, and we have confirmed these patterns hold when defining industry agglomeration through the Duranton and Overman (2005) index. These results suggest that entrepreneurship may be most important for industries that have the most powerful interactions among clustered firms. They also suggest that our results extend well beyond the growing demand of home markets. The last column shows a similar impact for highly agglomerated industries within trade, finance, and services.<sup>16</sup>

### 4.3 City-Industry Growth Regressions

While the correlation between entrepreneurship and urban employment growth for cities is quite strong and robust to covariates, our confidence in this link is also based upon its strength across industries within cities. Table 4 illustrates these connections. We define industries at the two-digit level of the Standard Industrial Classification system, and we continue to consider the metropolitan area in this analysis. To focus on meaningful variation, we require that industries have 100 employees throughout the period. This results in 12,178 observations. We continue to bootstrap standard errors.

Panels A and B again provide the results using average establishment size and start-up employment share, respectively. We refine our initial employment controls to be city-industry specific. We further include industry  $\times$  census division fixed effects in all specifications. These fixed effects account for the overall employment growth rate and entrepreneurship levels of each industry and region. The first column models the basic city growth covariates also used in Table 3. Columns 2-8 instead include city fixed effects that restrict variation to within-city differences. We thus look for connections of initial entrepreneurship to subsequent employment growth after removing overall patterns by city and by region-industry.

The correlation between our entrepreneurship measures and subsequent employment growth is typically smaller at the city-industry level. In the first column, we find that a one standard-deviation decrease in average establishment size is associated with a 0.19 standard-deviation increase in subsequent employment growth for the city-industry. A one standard-deviation increase in the share of employment in start-ups is associated with a 0.05 standard-deviation increase in subsequent employment growth. These effects are statistically significant and economically meaningful. The second column shows that these effects are only slightly diminished when we switch from city growth controls to city fixed effects.

These results suggest that the employment-entrepreneurship link is quite strong within cities, but that the effects are somewhat weaker than at the metropolitan area level. One explanation for the weakening of the effect is that perhaps entrepreneurship is proxying for other city-level

---

<sup>16</sup>Appendix Table 1 shows that the patterns hold when splitting our sample into two time periods. The results are stronger during 1982-1992 than during 1992-2002, although the differences between the two periods are not statistically distinct.

attributes. Another explanation is that there are cross-industry spillovers from entrepreneurship, as suggested by Chinitz’s hypothesis about a local culture of entrepreneurship.<sup>17</sup>

Columns 3-8 consider subsamples of the city-industry data; estimations include the most stringent city and industry x census division fixed effects. The first two columns again separate industry groups. The relationship between entrepreneurship and employment growth is robustly present in both groups, being stronger for mining, construction and manufacturing than for trade, services and finance. These results confirm our earlier findings for cross-metropolitan area employment growth, and they show power where the aggregate growth effect was weaker.<sup>18</sup> Columns 5 and 6 show similar results in cold and warm regions. Columns 7 and 8 find similar results by decade. Overall, these city-industry disaggregations and other unreported tests show the deep empirical association between initial entrepreneurship and subsequent growth. This association is more stable across decompositions at the city-industry level than at the city level.

## 5 Historical Mines and Modern Entrepreneurship

### 5.1 Historical Mines Data

While these patterns are provocative, the potential endogeneity of initial entrepreneurship remains worrisome. An abundance of start-ups in a particular city may reflect unmeasured city level attributes that make both entrepreneurship and future job growth more feasible. The concentration of entrepreneurship in particular city-industries could signal greater opportunities within that local economic sector or unobserved policy interventions. While the econometric tests reported above create a high bar for these alternative explanations, there is still a need to identify in this literature an exogenous source of variation in entrepreneurship. To address these issues, we now turn to the historical presence of mines close to each city.

We develop our instruments on the location of mines using several sources. Our primary data source on the geographic distribution of historical mines is the U.S. Geological Survey (USGS) database. This survey provides data on present and past mines, including their discovery dates and latitude-longitude spatial locations. We focus on mines that were known to exist in 1900. We believe that this survey provides a relatively complete survey of mineral and ore availability at the start of the 20th century. Deposits were a great source of wealth, and the government took its surveying responsibilities seriously. Congress established the USGS in 1879 and chose prominent early directors like Clarence King and John Wesley Powell to lead the organization. While it is possible that mineral and ore deposits were more likely to be discovered in areas

---

<sup>17</sup>Evidence for these cross-industry links have been identified in micro-data studies of the Chinitz effect like Rosenthal and Strange (2003, 2010), Glaeser and Kerr (2009), Glaeser, Kerr and Ponzetto (2010), and Drucker and Feser (2012). Hanlon (2012) and Helsley and Strange (2012) provide recent evidence on inter-industry linkages more broadly. Saxenian (1994), Davidsson (1995), Hofstede (2001), Lamoreaux, Levenstein and Sokoloff (2004), Landier (2006), and Falck, Fritsch and Heblich (2009) are examples of work on entrepreneurial culture.

<sup>18</sup>There is a subtle but important difference between the industry disaggregations in Tables 3 and 4. In Table 3, we maintain the same city-level entrepreneurship metrics to predict employment growth for both groups. In Table 4, the entrepreneurship measures are city-industry specific by definition.

that were more heavily inhabited or used for manufacturing during the 1800s, maps from the era certainly suggest that the USGS was doing a good job of surveying the entire country.<sup>19</sup>

The exact spatial locations of mines allow us to count mines that were known to exist in 1900 in spatial rings around cities. We design these spatial rings to be between 250 and 500 miles, and we provide below an analysis of price data from the time that leads us to these distance horizons. Our first instrument is the logarithm of the count of mines within 500 miles of the geographic centroid of the city in 1900. Cities have on average 943 mines in this spatial range, ranging from a minimum of ten to a maximum of 2966. We find very similar results to those reported below when weighting mine counts by the number of different types of ores that each mine extracts. We use the logarithm to allow for concavity in the impact of total mine counts. A few cities are not within 250 miles of a known mine in 1900. For this distance band, we add one to the count of mines before taking the logarithm.<sup>20</sup>

These initial instruments model the broad availability of natural deposits around cities, as mining and extractive industries broadly speaking are associated with larger establishment sizes. We complement this instrument with two additional metrics that describe the character of local deposits for the showcase example of the steel industry in the Chinitz hypothesis. Our first is an indicator variable for whether coal and iron ore is the dominant mining product of a state in 1928.<sup>21</sup> We take this measure from the 1930 Statistical Abstract of the United States, Table 739. We use this alternative source because the USGS data do not capture very well historical coal deposits, which were a very important spatial factor in industry location choice. Our final historical measure is the count of iron ore mines within 100 miles of the city in 1900. More than a third of cities do not have an iron ore mine within 100 miles, and we thus use the levels of this variable directly. The three different designs of the instruments (i.e., log count, indicator variables, mines count) also allow for capturing different aspects of the relationship.

## 5.2 Modern Mines Data

While the historical aspects of our data are important for introducing exogeneity to modern entrepreneurship, an alternative concern is that data quality is compromised by using information from the earlier period. The most important aspect of this liability for our current work is that the USGS data do not list the discovery date for most mines, and we have no way of assessing whether unreported dates are generally older or not (e.g., knowledge of the mine stretches so

---

<sup>19</sup>In the 1800s, prospecting often preceded industry, as it had, for example, in the California Gold Rush or the later Black Hills Gold Rush. Long before the upper peninsula of Michigan was well settled, the state government sent pioneering geologist Douglass Houghton to survey the area. Houghton would help establish the copper and iron ore deposits in the region. Likewise, a 1908 report already identifies the four largest coal deposits to be in Colorado, Montana, North Dakota and Wyoming, followed by West Virginia and Illinois, despite the fact that formal extraction at the time in Pennsylvania was an order of magnitude higher than any other state. See 1910 Statistical Abstract of the United States, Table 12, and 1930 Statistical Abstract of the United States, Table 767.

<sup>20</sup>These data are available and described at <http://tin.er.usgs.gov/mrds/about.php>. Appendix Tables 2a and 2b provide additional descriptive statistics on our mining data.

<sup>21</sup>States in this category are AL, CO, IL, IN, KY, MD, MI, MN, ND, PA, TN, VA, WA, and WV.

far back that a discovery date is unknown). Especially with instruments based upon natural resources, an argument can be made to utilize the raw capacity and inherent mineral wealth of a region, rather than knowledge of it at a particular point in history.

To address this issue, we report below additional results that use current information. For our two instruments developed from the USGS data, log count of total mines and local iron ore mine counts, we simply adjust the metric construction to build off all known mines in the database regardless of discovery date. For this purpose, we also develop a new instrument that utilizes the nature of coal deposits in a local area.

During the 1970s energy crisis, the USGS initiated a large-scale project to build a national coal information database that contains much deeper information about coal deposits throughout the country.<sup>22</sup> This database again includes latitude-longitude spatial locations, and it has a special feature that the types of coal are identified for mines. This is valuable information as coal deposits vary in grade and their spatial distribution. Anthracite coal, a particularly hard and compact form, is the most valuable but often quite difficult to supply. Bituminous coal, also known as black coal, is softer and less valuable than anthracite, but still widely mined, transported, and used in industrial applications. On the other hand, lignite coal, also known as brown coal, is of very low grade and often fails to be economical to mine and transport.

Figure 2 shows that these differences in coal type were known in 1900, but we do not have discovery dates that would facilitate instruments using coal grades circa-1900. We use this information, however, to create an alternative modern instrument that is an indicator variable for anthracite and bituminous coal being the predominant form of coal in a 150-mile spatial band around the city. The indicator variable takes a zero value if no modern coal deposits are within the band or if most deposits are lignite. Unlike our historical measure of whether coal and iron were the top state product in 1928, this modern instrument does not utilize realized production rates. We also use these data in two supplementary applications discussed next.

### 5.3 Selection of Spatial Rings

We now return to our selection of the spatial ring used for the total count of mines instrument. An important starting point is the identification that mineral deposits can influence cities over at least moderate spatial horizons. This reach descends in large part from the durable nature of minerals that aids in shipping them. By the early 20th century, transportation within the United States had reached a reasonable stage of development. Railroads and water transportation were strong by 1900 (e.g., Field 2011, Duran 2010), and the average price per ton-mile had declined from 6.2 cents in 1833 to 0.7 cents in 1900 (Carter et al., 2006). In the late 1800s, the cost of 10 miles of wagon transport was roughly equivalent to the cost of 375-475 miles of railroad or water transport, and the U.S. transportation network aided resource flows to cities beyond their immediate vicinity (Donaldson and Hornbeck, 2012). The relocation of some steel production

---

<sup>22</sup>These data are available and described at <http://energy.er.usgs.gov/products/databases/CoalQual/intro.htm>.

from Pittsburgh to Buffalo in the early 20th century reflected in part the ease of moving coal from Pennsylvania to New York and Buffalo's location on the shipping routes for iron coming from the west. These and related facts indicate that mines do not need to be immediately proximate to cities to influence their industrial structures.<sup>23</sup>

Unfortunately, while these basic concepts are known, the historical record for actual shipments of minerals and coal is very sparse and insufficient for detailed assessments. Our best evidence comes from coal price data across 47 cities in our sample for 1925-1930 reported in the 1940 Statistical Abstract of the United States, Table 772. This table separately lists prices of anthracite and bituminous coal. For most cities, prices are only given for a single type of coal reflecting that the city relied almost exclusively on that coal variant. We thus consider the price data in two ways. The first is a simple indicator variable by coal type for whether a price is given; the second is the log price of a coal variant conditional on a price being listed.

Tables 5a and 5b report results of regressions of these outcome variables on the spatial distributions of anthracite and bituminous coal deposits around each city, respectively. We utilize the modern coal database for these measures given the lack of historical records on coal variants. We report four distance horizons of 0-50, 50-100, 100-250, and 250-500 miles. The explanatory variable is the count of deposits within these bands, with counts normalized to have unit standard deviation for interpretation. We pool the data from all six years, clustering standard errors by city and including year fixed effects. We test with and without regional fixed effects; we find similar patterns if also controlling for water access to Great Lakes or the ocean. We have 261 observations where at least one price is listed, 133 where an anthracite price is listed, and 216 where a bituminous price is listed.

In Table 5a, we find that mines up to 250 miles distance from a city are important for explaining whether anthracite coal was in use and its price level. On the other hand, anthracite mines from 250-500 miles only exhibit a strong association for log prices when controlling for region effects. In Table 5b, there is not a clear pattern for whether a bituminous price is listed in Columns 1 and 2. On the other hand, Columns 3 and 4 find a strong association for regional deposits of 100-500 miles lowering bituminous coal prices in the cities.

Our assessment from these various data points is that the spatial band for total mine counts should be at least 250 miles. The above price rings are built off of coal, which is a heavy product compared to many other minerals. Thus, the fact that the deposit influence is evident to 500 miles for coal prices suggests that this spatial range is likely to be true for many other minerals. We thus test below setting the bands for total mine counts at 250 and 500 miles.

---

<sup>23</sup>The economic history accounts of whether natural advantages or market access determined the spatial placement of large-scale manufacturing by 1900 are mixed. See Krugman (1991), Kim (1995), and Klein and Crafts (2009). Related work on industry location and natural advantages includes Ellison and Glaeser (1999), Kim (1999), Rosenthal and Strange (2001, 2004), Glaeser and Kerr (2009), Combes et al. (2010), Holmes and Lee (2012), Ellison, Glaeser and Kerr (2010), Kerr and Kominers (2010), and Storeygard (2012). Our work is closely related to the path dependency study of Bleakley and Linn (2012) around historical portage sites. Dippel (2012) considers historical mines placements in the context of Native American integration and economic development.

As our estimations include fixed effects for the nine census divisions, we only identify off of city differences in proximity to historical mining deposits within each region. Levels differences across the nine census divisions account for about a quarter of the total variation across cities at 500 miles. This regional explanatory power is similar when using a 100 or 250 mile radius.

## 5.4 Historical Mines and Modern Entrepreneurship

Table 6 shows that our mining metric strongly predicts entrepreneurship late in the 20th century. Column headers indicate outcome variables, and the regressions also control for census division fixed effects, initial employment, and city growth covariates. Panel A reports estimates with the log count of mines within 500 miles as the central explanatory variable. As the covariates are the same variables that will be included in our final regressions, Columns 3 and 6 thus represent first-stage relationships.

The first regression in Panel A shows the connection between the number of mines and the average establishment size in manufacturing in 1963. We do not have data for a wider range of industries during that year. As the number of mines increases by one standard deviation, the average establishment size in manufacturing increases by 0.21 standard deviations. This relationship is both statistically significant and economically relevant. The t-statistic is about three. We have also confirmed that mines in 1900 are associated with weaker entrepreneurship for manufacturing in the 1960s.

Column 2 shows the strong relationship between historical mines and mining activity at the start of our time period. A one standard-deviation increase in the number of mines is associated with a 1.16 standard-deviation increase in mining employment near the city over 1976-1980. These deposits certainly still matter for the industrial composition of an area.

Column 3 looks at the relationship between historical mining deposits and average establishment size in 1982, the relevant year for our instrumental variables estimations. The estimated elasticity is 0.075, which means that as the number of mines increases by one standard deviation, average establishment size increases by about 0.08 standard deviations. The t-statistic of this effect is more than six. Unreported regressions find that the similar effect for 1992 weakens by about a quarter but remains quite significant.

The fourth and fifth columns show the relationship to average establishment size in the two sectors. The estimated elasticity is three times higher in mining, construction and manufacturing than in trade, finance and services. A one log point increase in the number of mines raises average establishment size in closely related sectors by more than ten percent and in unrelated sectors by four percent. Both estimates are statistically significant. The final regression shows that historical mining deposits are also predictive of the city's start-up employment share in 1982. The overall elasticity estimate is -0.16.

Panel B extends the estimation in Panel A to also include an indicator variable for whether coal and iron ore was the top mineral product of the state. This starts to model the types of

mines that surround a city. This indicator variable is also very predictive of increases in average establishment size and reduced entry rates. This suggests that coal and iron ore deposits are especially important for large-scale operations conditional on the number of mines surrounding a city.

Panel C reports results using the log count of mines with 250 miles by itself. The elasticities at this spatial level are about half of those using the 500-mile spatial bands, and the coefficients are more precisely estimated. The most substantive change is the weaker link of mines to establishment size in trade, finance and services. Panel D alternatively reports results by two distance rings of 0-100 and 100-500 miles estimated jointly. As more than a quarter of cities do not have a mine within 100 miles, we use a levels regression that allows for zero values. Coefficients and standard errors are multiplied by 100 for visual clarity. For most of the outcome variables, the presence of mines within 100 miles matters two- and threefold more than mines over 100-500 miles.<sup>24</sup> On the other hand, similar to Panel C, the very localized presence of mines does not predict average establishment size in unrelated sectors of trade, finance and services. This effect comes mostly through mines in the larger spatial area around the city.

Finally, Panel E examines concentrations of anthracite/bituminous deposits using current data. There are visible connections between coal grade composition, mining sector development, and modern establishment size. In another test, we regress the average establishment size of a city in 1982 on the count of anthracite/bituminous deposits within 150 miles, the count of lignite deposits within 150 miles, and our standard covariates. A one standard-deviation increase in anthracite/bituminous deposits is associated with a 0.030 (0.006) increase in log average establishment size, while the elasticity for lignite is 0.007 (0.007). The elasticities are similarly 0.029 (0.006) and 0.007 (0.008) when using each mine type individually. This test, while admittedly crude, confirms that the nature of deposits is important for our assessment. It also provides some confidence that the use of minerals is important, rather than spurious features of the geographic landscape (e.g., rugged mountain terrain).

These regressions ensure that the problem with our instruments will typically not be in their first-stage fit. Mines in 1900 are strongly related to establishment size and entrepreneurship at the beginning of our regression time period. Our larger concern is that mines could easily be correlated with employment growth for reasons other than initial entrepreneurship. We will address this concern after presenting our core instrumental variables results.

---

<sup>24</sup>These patterns also hold when using more disaggregated bands, suggesting mostly regular declines in the impact of mines on industrial structures with greater distance. When using three distance bands of 0-100 miles, 100-250 miles, and 250-500 miles, the coefficients for average establishment size are 0.016, 0.022, and 0.009, respectively. Those for birth shares are -0.075, -0.045, and -0.024. All estimates are statistically significant.

## 6 Instrumental Variables Results

### 6.1 City Growth Estimations

Table 7 describes our key second-stage results of entrepreneurship and local growth using proximity to mines in 1900 as instruments. Panel A considers average establishment size in 1982 as the core independent variable, while Panel B models initial entrepreneurship through the local employment share in start-ups. Regressions control for census division fixed effects, initial employment, and city growth covariates. We report bootstrapped standard errors.<sup>25</sup>

Column 1 begins with a single instrumental variable regression using the log count of mines in 1900 as the instrument, finding that the effect of average establishment size on subsequent growth increases substantially when using mines as an instrument. The relevant ordinary least squares coefficient is -0.69, and this instrumental variables estimate is -0.97, which means that a standard-deviation increase in a city’s average establishment size is associated with a standard-deviation decrease in employment growth over 1982-2002. For Panel B’s employment share in start-ups, the coefficient increases from 0.16 to 0.45. Both estimates have t-statistics greater than 2.5. The associated diagnostic tests indicate that the instrument performs well for the full sample.

Column 2 adds a second instrument of the indicator variable for dominant product type, and Column 3 further expands to the triple instrument specification that also includes the count of iron ore mines with 100 miles as an instrument. The additional instruments modestly reduce the coefficients and sharpen the precision of the estimates. These results suggest instrumented elasticities of about -0.9 for average establishment size and 0.4 for start-up employment shares, respectively. The various diagnostic tests continue to perform well, with the one exception that the over-identification test for the triple instrument in Panel B is rejected at a 10% level. While differences shrink when using multiple instruments, it is still the case that the measured elasticities are higher than in ordinary least squares.

Columns 4 and 5 repeat Columns 1 and 3, respectively, using the 250-mile spatial band rather than the 500-mile spatial band. The impact of this change is to lower the estimated second-stage elasticities to be comparable to ordinary least squares estimates. The instrumented effect of average establishment size is -0.52 to -0.61, smaller than the ordinary least squares coefficient of -0.69, while it is 0.25 for start-up employment, larger than the ordinary least squares coefficient of 0.16. Tests do not reject that these coefficients are the same.

Combining these approaches, Column 6 reports results using four instruments that include both 250- and 500-mile spatial bands. These results sit in-between those of Columns 3 and 6. Going forward, we report our results using the two bands individually as they bound this joint

---

<sup>25</sup>Similar to least squares, we find smaller standard errors when clustering our instrumental variable regressions by region. Bester et al. (2011) demonstrate how clustering by large, contiguous groups of approximately similar size with substantial interiors relative to boundaries can appropriately model spatial decay dependency. We also find smaller standard errors when using spatial decay frameworks like Drukker et al. (2011).

effect. We view the 500-mile band as making the maximum case for entrepreneurship's role, and the 250-mile band as making the minimum case based upon historical mines. Finally, Column 7 shows very similar results when using instruments based upon modern data.<sup>26</sup>

The overall patterns from Table 7 suggest that instrumental variables estimates are comparable to or higher than ordinary least squares estimates. What can account for this feature? A first, relatively mundane, explanation is that the instrumental variables are correcting for measurement error in the regressors that downward biases ordinary least squares estimates. Our regressors are measured at a point in time at the start of the sample period, and thus they may be sensitive to idiosyncratic blips in city features. The employment share in start-ups seems the more exposed metric to this issue, and this perhaps explains why its relative increases in instrumented elasticities compared to ordinary least squares estimates are stronger than those for average establishment size.

A second explanation is that the endogenous aspects of average establishment size and new start-ups actually work against city growth, while the exogenous aspects—captured by the long-run supply of entrepreneurs—have an even stronger positive effect than the ordinary least squares estimates indicate. According to this view, negative aspects of an area kill off large firms and employment in older establishments, making average establishment size smaller and the start-up share larger. This is particularly important if urban decline pushes displaced workers into sub-optimal entrepreneurship that is not growth enhancing. By allowing only the variation that comes from the long-run supply of entrepreneurs to influence our estimates, the instrumental variables estimates correctly show a larger elasticity of long-run growth with respect to entrepreneurship.

A less positive, third interpretation is that mines are positively associated with other aspects of the city that are connected with longer term decline. According to this view, the orthogonality condition needed for the instrumental variables estimation is violated by a correlation with omitted variables and this correlation causes the instrumental variables estimates to be artificially high. The over-identification tests are one econometric assessment of this concern, and our key results usually pass these tests. We further focus the rest of this paper on this potential problem using sample decomposition and quantile instrumental variable techniques.

Before starting with the sample decompositions, we explicitly test one alternative story. Holmes (2006) finds a very striking connection between local dependence on mines and unionism. Similar to our analysis, Holmes notes the extent to which unionism "spills out coal mines and steel mills into other establishments in the neighborhood, like hospitals and supermarkets." The analysis identified the potential channels of a common local infrastructure for unionism and contagious attitudes among families and friends toward labor organization.<sup>27</sup> To ensure that

---

<sup>26</sup>To conserve space, we only report employment results for the instrument variable specifications. We continue to find that employment and payroll growth closely track each other. Disaggregating the 1982-2002 employment growth into five-year intervals, growth effects are evident in each sub-interval except 1992-1997. We also find similar results using LIML estimators.

<sup>27</sup>We thank Curtis Simon for sharing this lyric: "My daddy was a miner, And I'm a miner's son, And I'll stick

unionism is not driving our results, we develop from Hirsch and Macpherson (2003) estimates of 1982 union membership rates for 214 cities in our sample. Across these cities, our base instrumented elasticity is -0.594 (0.326). This elasticity ranges between -0.600 (0.296) and -0.525 (0.346) after including the union control depending upon how it is entered. Thus, while unionism and entrepreneurship are surely connected and both influenced by historical mining legacies, this alternative channel does not appear to be solely driving our results.

## 6.2 Sample Decomposition

Appendix Tables 3a and 3b provide decompositions of effects using the 500-mile band instruments. Appendix Tables 4a and 4b provide similar results for the 250-mile band instruments. In both cases, the table format mirrors that used in Table 3 with a total effect and then allowing a difference in treatment effects between warm and cold places. For convenience, Figure 3 plots the ordinary least squares effects and instrumental variable effects using both distance bands for average establishment size. Likewise, Figure 4 plots the various effects for the start-up entry shares.

Examining Figure 3, a first observation is that the general patterns evident in Table 7 persist between the two distance bands. Using the 500-mile band leads to larger effects than least squares that are statistically different from zero with all of our different decompositions. On the other hand, the 250-mile band estimates more closely mirror the least squares results. The effects are statistically different from zero for trade, finance, and services sectors and for industries with moderate-to-high levels of agglomeration. On the other hand, the effect is not statistically significant for mining, construction, and manufacturing.

This weaker performance for mining, construction, and manufacturing compared to trade, finance, and services sectors is quite intriguing. The former is the part of the economy where we would think that the direct effect of mines is likely to be most severe, while the latter is less prone to a direct effect of mines on growth. These results suggest to us that omitted variables related to sector demand declines are not driving the results. While it is certainly reasonable that declines in manufacturing or mining sectors that are tightly connected to historical mines would also depress local employment in other industries due to weak demand, it is hard to believe that this demand-side spillover effect would be larger for those other industries than for mining itself.

Likewise, the variation across industries by their level of agglomeration is insightful as spatial industrial concentration is one measure of the extent to which an industry is focused on supplying the local market. Industries that focus on supplying local customers (e.g., barbers, restaurants) tend to be ubiquitous and therefore non-agglomerated. On the other hand, industries that focus on serving a global market have less reason to spread themselves out and therefore tend to be more agglomerated (e.g., movie production, automobile manufacturers, investment bankers).

---

with the union, Till every battle is won." from "Which Side are You On?" by Florence Reese.

The effects we find are most pronounced in agglomerated sectors.

This logic pushes us to focus on the most highly agglomerated industries within the trade, services, and finance sector. These agglomerated industries seem least likely to be directly influenced by any decline in local manufacturing or mining associated with the direct effect of mines. The results here depend on the spatial band. With a 500-mile band, we continue to find a strong role for entrepreneurship and local growth in these specialized sectors. With the 250-mile band, we do not find a strong relationship. This difference relates back to Table 6’s finding that moderately distant mines appear to have more important effects for the trade, services, and finance sector than very proximate mines.<sup>28</sup>

Figure 4 extends this framework to the start-up entry shares. The basic patterns across sectors mirror those of Figure 3. One continues to find substantial and statistically significant effects with the 500-mile band instruments, while the 250-mile band instruments are only statistically significant in the case of highly agglomerated industries.

Within Appendix Tables 3a-4b, Panel B presents results where we allow for effects to differ by warm and cold regions. We undertake this split as we suspect *ex ante* that the omitted variables correlations are most severe in colder areas where industrial decline has been acute. As this logic suggests, the estimated coefficients are generally higher in the colder regions than in warmer cities, where manufacturing decline has been far less pronounced. The more substantive message from this decomposition, however, is that differences between warm and cold areas are fairly small. This stability provides comfort that our results are not simply following from the decline of the Rust Belt or similar.<sup>29</sup>

### 6.3 City Growth Projections

The D panels of Figures 3 and 4 show our primary employment growth results when including dynamic controls. Appendix Tables 5a and 5b provide the underlying estimations. These controls model the projected path of the city during the 1982 to 2002 period due to the city’s industry composition (e.g., is the city home to industries generally growing or contracting).

We first test including the project forward employment growth of the city due to its industry composition. We calculate the projected forward growth by interacting the initial 1982 industry distribution of the city at the two-digit SIC level with the employment growth nationally of the industry from 1982 to 2002 outside of the focal city. Introducing this control lowers the role of initial entrepreneurship somewhat. On the other hand, introducing a similar control based

---

<sup>28</sup>We also find a similar emphasis when linking average establishment size by sector with that sector’s employment growth. These results again highlight that most of the growth effects that we are capturing come outside of sectors traditionally dependent upon mines. While we believe that average establishment size across the whole city is the more appropriate metric, it is comforting to find similar patterns when focusing just on the trade, finance and services sectors.

<sup>29</sup>In broader terms, our growth covariates capture many features that are known to increase city attractiveness and urban growth. We find similar results when using hedonic housing price regressions similar to Glaeser et al. (2010) to measure amenities of cities.

upon establishments count distributions and their forward projection raises the role of initial entrepreneurship somewhat.

We next test a variation of the above approach. It could be that what matters more for a city than its own predicted path is its relative rank to other nearby cities. To test this, we develop 18 cells that are the interaction of the nine Census regions with whether or not a city has an above-average growth projection in 1982. We then introduce fixed effects for these cells so that we compare amongst cities within a region with similar growth projections. This approach yields very similar results to our primary estimations.

Finally, we have manufacturing data that goes back to the 1963 Census of Manufacturers. We include a control for the log growth in manufacturing employment or establishments for the city from 1963 to 1981. This again has little effect on our estimated coefficients. Overall, these results suggest that the link of initial entrepreneurship to employment growth is not reflecting simple trend differences in city growth prospects due to industry composition.

## 6.4 Instrumental Variable Quantile Regressions

We complement the decomposition exercises, which focus on differentiating treatment effects by fixed traits of cities or industries, with a second analysis that considers whether the effects of entrepreneurship vary between rapidly and slowly growing cities. Because this approach conditions on the growth outcome variable, rather than a fixed trait, it requires a more careful procedure that combines causal inference and heterogeneous treatment effects. The instrumental variable quantile regression method (IVQR) of Chernozhukov and Hansen (2004a, 2005, 2006) proves very suitable for our current empirical setting. We provide here a qualitative introduction, with the appendix and referenced papers providing a technical description.

One can think of IVQR as beginning with a determination of the conditional growth rate of a city. The conditional aspect means that one has controlled for important covariates that systematically determine whether cities are growing fast or slowly, so that one is looking at the unexplained growth above and beyond these basic traits.<sup>30</sup> The procedure then estimates local instrumental variable treatment effects for various points in the conditional distribution by weighting nearby points of the distribution more than distant points. Thus, when estimating the instrumental variable elasticity for the 25th quantile of the growth distribution, the procedure places greater emphasis on the empirical links between mines, entrepreneurship, and growth among other slow growing cities. The experiences of fast growing cities, on the other hand, would receive more weight when estimating effects at the 75th quantile.

Figures 5 and 6 providing graphical depictions of the IVQR results, with exact values for

---

<sup>30</sup>This methodology has a demanding assumption of rank invariance (or similarity) that makes the inclusion and correct specification of the covariates very important. Rank invariance requires that treatment status not disturb the cities' underlying ranks in the conditional growth distribution. Aspects of spatial distribution of urban growth—like the very strong regularities for faster growth of warm cities or skilled cities over the last few decades—are amenable for this setting. The R2 squared value of city growth on the initial growth covariates is 0.43.

selected quantiles given in Appendix Table 6. Starting with average establishment size in Figure 5, the quantile regression plot in Panel A shows remarkable homogeneity across the conditional growth distribution in how entrepreneurship connects with city growth. The lower quantiles on the left hand side of each graph indicate entrepreneurship’s role among cities that are growing slower than their regional location, climate, 1970s housing prices, and so on would predict. Those on the right are growing faster. The differences appear quite limited, and Wald tests do not reject that the coefficients are the same at the 15th and 85th percentiles of the distribution. The instrumented elasticities when using the 500-mile bands also display homogeneity and are always statistically significant. When using the 250-mile bands, statistically significant effects are mostly evident in quantiles up to the median of the distribution, but not in very fast growing cities. Figure 6’s depiction of the impact of start-up employment leads to similar conclusions.

The homogeneity of our effects across the conditional growth distribution that this procedure identifies is very important. Our central concern has been that the historical presence of mines depressed city growth due to factors unrelated to reduced modern entrepreneurship. These results suggest that our overall treatment effects in Table 7 are not coming from abnormalities in one part of the growth distribution, but instead are much more broad-based. If anything, we find that entrepreneurship’s role may be more important for employment growth in cities that are underperforming expectations, but the differences are not statistically significant.

## 7 Extended Employment Growth Results

This section provides several extensions to our work to further clarify entrepreneurship’s role in city growth. Table 8a reports these analyses using average establishment size, and Table 8b considers start-up employment shares. In both tables, Panel A presents least squares results, Panel B presents results using the triple instruments and the 500-mile band, and Panel C presents results using the triple instruments and the 250-mile band.

The first column tests a redefinition of our employment growth variable. Thus far, we have considered total city growth from 1982 to 2002. Column 1 makes use of the micro-data to identify the employment in 2002 of establishments born since 1982. We then measure the log ratio of the net employment generated in new establishments to the initial city size in 1982. This measure thus removes any growth associated with incumbent enterprises in 1982. The elasticities with this measure are quite strong and robust across the instrument designs. This exercise identifies the higher direct employment contribution from the new entrants.

Columns 2-4 take a second perspective. We now calculate the share of employment overall and broken down by establishment size in 2002 in the city that is from entrants born after 1982. By considering shares of activity in 2002, we are no longer considering the growth of city employment itself but instead the composition of establishments in 2002. Column 2 of both tables shows that cities with greater entrepreneurship in 1982 have a larger share of their 2002 employment

contained in new enterprises. In the instrumented regression, a 10% increase in 1982 average establishment size is associated with a 1.4% decrease in the city’s new-entrant employment share in 2002; similarly, a 10% increase in initial birth employment shares is associated with a 1% increase in the city’s new-entrant employment share. Columns 3 and 4 partition this effect by 2002 establishment size, finding that higher initial entrepreneurship especially connects to a greater new-entrant share among establishments with more than 100 employees in 2002.

As an alternative, Column 5 models the average age of establishments in 2002 for the city. We calculate ages from the LBD’s start in 1976, giving a maximum of 26 years old. We weight establishment ages by the 2002 employment in establishments (results are very similar unweighted). In the least squares framework, a 10% increase in average initial establishment size is associated with a 2% older age profile in 2002; a higher birth employment share predicts in a similar way a lower average age for 2002. Column 6 shows that this younger age effect is present when isolating just the unweighted average age of the top 25 employers for 2002 in each city. These age effects is even sharper when instrumenting using historical mines placements. Thus, higher initial entrepreneurship of the city is associated with a younger establishment age profile of the city in 2002, even for top employers.<sup>31</sup>

On the whole, these patterns support the primary link established for initial entrepreneurship and city growth. Expanded employment is generated in new establishments, and cities with higher initial entrepreneurship show lower age structures, even among their top employers, two decades later. Haltiwanger et al. (2012) describe an important up-or-out dynamic that connects young firms to firm-level employment growth. These patterns suggest a similar process is occurring at the city level, with successful start-ups expanding to become larger establishments and thereby generating employment growth. Evidence of these dynamics sit more closely with industrial legacies like the Chinitz hypothesis and entrepreneurship’s role than if, for example, the employment growth came solely through older incumbents or endless replications of very small firms.

## 8 Conclusion

The correlation between measures of entrepreneurship—such as the share of local employment in new start-ups or the average establishment size—and subsequent urban employment growth is quite robust both across and within cities. One concern with these measures is that they may capture other aspects of the local environment besides entrepreneurship. This paper tried to push forward on these issues by looking for the historical roots of small establishment sizes and

---

<sup>31</sup>Unreported regressions analyze the forward evolution of incumbent firms in 1982 as a function of their local initial entrepreneurship. The least squares and instrument results both find that survival prospects for 1982 incumbents are decreasing in initial entrepreneurship for the city. Least squares finds incumbent establishments that survive in places with higher initial entrepreneurship tend to also grow more, but the instrumental variable regressions do not support these results. The localized link of entry and exit rates is observed, for example, by Dunne et al. (1988), Davis et al. (1996), and Kerr and Nanda (2009).

higher entry rates.

We followed the intuition of Chinitz (1961), who argues that industries dependent upon mineral and coal deposits, like steel, involve large companies that create executives, not entrepreneurs. We use the presence of mineral and coal deposits in 1900 to provide us with variation in the level of resource-intensive industries. These deposits are associated with larger establishment sizes and lower birth employment shares in the 1960s and onwards. Using this spatial proximity for instruments, we continue to find a significant link between our measures of entrepreneurship and urban employment growth.

The big concern with this variable is that it is quite plausibly correlated with aspects of the local economy other than entrepreneurship, such as manufacturing decline. We tried to control for these factors with city-level variables, region fixed effects, and so on, but we recognize that our measures are far from perfect. We focused then on industries that were not directly related to mining, and on industries that were highly concentrated spatially, which suggests that they do not depend on a local market. We also focused on warmer cities, which should be less sensitive to the decline of the Rust Belt, and we modeled city growth projections. Our core results remain unchanged. Finally, using quantile instrument variable techniques, we identified that our effects are present in both cities exceeding and underperforming growth expectations based upon their initial traits.

The weight of this evidence suggests that entrepreneurship is playing an important role in modern urban growth. While we have tried to systematically address concerns about the correlation between our instrument and the error term, we remain overall cautious about our results. We hope that our work prompts other researchers to identify sources of exogenous variation in urban entrepreneurship, within the United States or outside of it. The patterns in Figure 1 are exceptionally strong and the backbone for many policy initiatives. It is remarkable how little we know about what lies behind this relationship, especially given how widely-held the belief is that entrepreneurship is important for economic performance.

## References

- [1] Acs, Zoltan, and Catherine Armington, *Entrepreneurship, Geography and American Economic Growth* (New York, NY: Cambridge University Press, 2006).
- [2] Agrawal, Ajay, Iain Cockburn, Alberto Galasso, and Alexander Oettl, "Why are Some Regions More Innovative than Others? The Role of Firm Size Diversity", NBER Working Paper 17793 (2012).
- [3] Bester, C. Alan, Timothy Conley, and Christian Hansen, "Inference with Dependent Data using Cluster Covariance Estimators", *Journal of Econometrics* 165:2 (2011), 137-151.
- [4] Blau, Peter, and Otis Dudley Duncan, *The American Occupational Structure* (New York, NY: Wiley, 1967).
- [5] Bleakley, Hoyt, and Jeffrey Lin, "Portage and Path Dependence", *Quarterly Journal of Economics* 127 (2012), 587-644.
- [6] Carter, Susan, Scott Gartner, Michael Haines, Alan Olmstead, Richard Sutch, and Gavin Wright, *Historical Statistics of the United States Volume 4* (Cambridge: Cambridge University Press, 2006).
- [7] Chernozhukov, Victor, and Christian Hansen, "The Effects of 401(k) Participation on the Wealth Distribution: An Instrumental Variables Quantile Regression Analysis", *Review of Economics and Statistics* 86 (2004a), 735-751.
- [8] Chernozhukov, Victor, and Christian Hansen, "Instrumental Variables Quantile Regression", MIT Department of Economics Working Paper (2004b).
- [9] Chernozhukov, Victor, and Christian Hansen, "An IV Model of Quantile Treatment Effects", *Econometrica* 73:1 (2005), 245-261.
- [10] Chernozhukov, Victor, and Christian Hansen, "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models", *Journal of Econometrics* 132:2 (2006), 491-525.
- [11] Chinitz, Benjamin, "Contrasts in Agglomeration: New York and Pittsburgh", *American Economic Review Papers and Proceedings* 51:2 (1961), 279-289.
- [12] Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux, "Estimating Agglomeration Economies with History, Geology, and Worker Effects" in Glaeser, Edward (ed.) *Agglomeration Economics* (Chicago, IL: University of Chicago Press, 2010).
- [13] Dahl, Michael, and Olav Sorenson, "Home Sweet Home: Social Capital and Location Choice", Working Paper (2007).
- [14] Davidsson, Per, "Culture, Structure and Regional Levels of Entrepreneurship", *Entrepreneurship & Regional Development* 7:1 (1995), 41-62.
- [15] Davis, Steven, John Haltiwanger, and Scott Schuh, *Job Creation and Destruction* (Cambridge, MA: MIT Press, 1996).
- [16] Day, David, *Mineral Resources of the United States: Calendar Year 1900* (Washington, DC: Government Printing Office, 1901).

- [17] Delgado, Mercedes, Michael Porter, and Scott Stern, "Clusters and Entrepreneurship", *Journal of Economic Geography* 10:4 (2010a), 495-518.
- [18] Delgado, Mercedes, Michael Porter, and Scott Stern, "Clusters, Convergence and Economic Performance", CES Working Paper 10-34 (2010b).
- [19] Dippel, Christian, "Forced Coexistence and Economic Development: Evidence from Native American Reservations", Working Paper (2012).
- [20] Donaldson, Dale, and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach", Working Paper (2012).
- [21] Drucker, Joshua, and Edward Feser, "Regional Industrial Structure and Agglomeration Economies: An Analysis of Productivity in Three Manufacturing Industries", *Regional Science and Urban Economics* 42:1-2 (2012), 1-14.
- [22] Drukker, David, Ingmar Prucha, and Rafal Raciborski, "A Command for Estimating Spatial-Autoregressive Models with Spatial-Autoregressive Disturbances and Additional Endogenous Variables", *The Stata Journal* 1:1 (2011), 1-13.
- [23] Dunne, Timothy, Mark Roberts, and Larry Samuelson, "Patterns of Firm Exit and Entry in U.S. Manufacturing Industries", *Rand Journal of Economics* 19 (1988), 495-515.
- [24] Duran, Xavier, "A Model of Formation of Profit Expectations of Theodore Judah and the Expected Private Profitability of the First Transcontinental Railroad", Working Paper (2010).
- [25] Duranton, Gilles, and Henry Overman, "Testing for Localization Using Micro-Geographic Data", *Review of Economic Studies* 72 (2005), 1077-1106.
- [26] Elfenbein, Daniel, Barton Hamilton, and Todd Zenger, "The Small Firm Effect and the Entrepreneurial Spawning of Scientists and Engineers", *Management Science* 56:4 (2010), 659-681.
- [27] Ellison, Glenn, and Edward Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy* 105 (1997), 889-927.
- [28] Ellison, Glenn, and Edward Glaeser, "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?", *American Economic Review Papers and Proceedings* 89 (1999), 311-316.
- [29] Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", *American Economic Review* 100:2 (2010), 1195-1213.
- [30] Falck, Oliver, Michael Fritsch, and Stephan Heblich, "Bohemians, Human Capital, and Regional Economic Growth", Jena Economic Research Papers 2009-049 (2009).
- [31] Field, Alexander, "The Adversity/Hysteresis Effect: Depression Era Productivity Growth in the U.S. Railroad Sector", in Lerner, Josh, and Scott Stern (eds.), *NBER Rate and Direction of Inventive Activity 50th Anniversary Conference Volume* (2011).
- [32] Figueiredo, Octávio, Paulo Guimaraes, and Douglas Woodward, "Home-Field Advantage: Location Decisions of Portuguese Entrepreneurs", *Journal of Urban Economics* 52:2 (2002), 341-361.

- [33] Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, "Human Capital and Regional Development", *Quarterly Journal of Economics* (2012), forthcoming.
- [34] Ghani, Ejaz, William Kerr, and Stephen O'Connell, "South Asia—Entrepreneurship, Growth, and Job Creation", in Ghani, Ejaz (ed.), *Reshaping Tomorrow: Positioning South Asia for the Big Leap* (Oxford, UK: Oxford University Press, 2011).
- [35] Glaeser, Edward, and Joshua Gottlieb, "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States", *Journal of Economic Literature* 47:4 (2009), 983-1028.
- [36] Glaeser, Edward, Hedi Kallal, José Scheinkman, and Andrei Shleifer, "Growth in Cities", *Journal of Political Economy* 100:6 (1992), 1126-1152.
- [37] Glaeser, Edward, and William Kerr, "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?" *Journal of Economics & Management Strategy* 18:3 (2009), 623-663.
- [38] Glaeser, Edward, William Kerr, and Giacomo Ponzetto, "Clusters of Entrepreneurship", *Journal of Urban Economics* 67:1 (2010), 150-168.
- [39] Glaeser, Edward, José Scheinkman, and Andrei Shleifer, "Economic Growth in a Cross-section of Cities", *Journal of Monetary Economics* 36:1 (1995), 117-143.
- [40] Gompers, Paul, Josh Lerner, and David Scharfstein, "Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999", *Journal of Finance* 60:2 (2005), 577-614.
- [41] Haines, Michael, "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002", ICPSR 2896 (2005).
- [42] Haltiwanger, John, Ron Jarmin, and Javier Miranda, "Who Creates Jobs? Small vs. Large vs. Young", *Review of Economics and Statistics* (2012), forthcoming.
- [43] Hanlon, Walker, "Industry Connections and the Geographic Location of Economic Activity", Working Paper (2012).
- [44] Helsley, Robert, and William Strange, "Coagglomeration", Working Paper (2012).
- [45] Hirsch, Barry, and David Macpherson, "Union Membership and Coverage Database from the Current Population Survey: Note", *Industrial and Labor Relations Review* 56:2 (2003), 349-354.
- [46] Hofstede, Geert, *Culture and Organizations* (London, UK: Harper Collins, 2001).
- [47] Holmes, Thomas, "Geographic Spillover of Unionism", FRB Minneapolis Working Paper 368 (2006).
- [48] Holmes, Thomas, and Sanghoon Lee, "Economies of Density versus Natural Advantage: Crop Choice on the Back Forty", *Review of Economics and Statistics* 94:1 (2012), 1-19.
- [49] Hout, Michael, and Harvey Rosen, "Self-Employment, Family Background, and Race", *Journal of Human Resources* 35:4 (2000), 670-692.

- [50] Hurst, Eric, and Ben Pugsley, "What Do Small Businesses Do?", *Brookings Papers on Economic Activity* (2012), forthcoming.
- [51] Jacobs, Jane, *The Economy of Cities* (New York, NY: Vintage Books, 1970).
- [52] Jarmin, Ron, and Javier Miranda, "The Longitudinal Business Database", CES Working Paper (2002).
- [53] Kerr, William, and Scott Kominers, "Agglomerative Forces and Cluster Shapes", NBER Working Paper 16639 (2010).
- [54] Kerr, William, and Ramana Nanda, "Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship", *Journal of Financial Economics* 94 (2009), 124-149.
- [55] Klepper, Steven, and Sally Sleeper, "Entry by Spinoffs", *Management Science* 51:8 (2005), 1291-1306.
- [56] Kim, Sukkoo, "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U. S. Regional Manufacturing Structure, 1860-1987", *Quarterly Journal of Economics* 110:4 (1995), 881-908.
- [57] Kim, Sukkoo, "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987", *Regional Science and Urban Economics* 29:1 (1999), 123-137.
- [58] Klein, Alexander, and Nicholas Crafts, "Making Sense of the Manufacturing Belt: Determinants of U.S. Industrial Location, 1880-1920", Working Paper (2009).
- [59] Koenker, Roger, and Gilbert Bassett, "Regression Quantiles", *Econometrica* 46:1 (1978), 33-50.
- [60] Kortum, Samuel, and Josh Lerner, "Assessing the Contribution of Venture Capital to Innovation", *RAND Journal of Economics* 31:4 (2000), 674-692.
- [61] Krugman, Paul, "History and Industry Location: The Case of the Manufacturing Belt", *American Economic Review Papers and Proceedings* 81:2 (1991), 80-83.
- [62] Lamoreaux, Naomi, Margaret Levenstein, and Kenneth Sokoloff, "Financing Invention During the Second Industrial Revolution: Cleveland, Ohio, 1870-1920", NBER Working Paper 10923 (2004).
- [63] Landier, Augustin, "Entrepreneurship and the Stigma of Failure", Working Paper (2006).
- [64] Marshall, Alfred, *Principles of Economics* (London, UK: MacMillan and Co., 1920).
- [65] Michelacci, Claudio, and Olmo Silva, "Why So Many Local Entrepreneurs?", *Review of Economics and Statistics* 89:4 (2007), 615-633.
- [66] Miracky, W., "Economic Growth and Business Cycles in Cities: The Role of Locational Externalities", Thesis, Massachusetts Institute of Technology, 1993.
- [67] Niittykangas, Hannu, and Hannu Tervo, "Spatial Variations in Intergenerational Transmission of Self-Employment", *Regional Studies* 39:3 (2005), 319-332.

- [68] Rosenthal, Stuart, and William Strange, "Evidence on the Nature and Sources of Agglomeration Economies", in J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics, Volume 4* (Amsterdam: North-Holland, 2004), 2119-2171.
- [69] Rosenthal, Stuart, and William Strange, "Geography, Industrial Organization, and Agglomeration", *Review of Economics and Statistics* 85:2 (2003), 377-393.
- [70] Rosenthal, Stuart, and William Strange, "Small Establishments/Big Effects: Agglomeration, Industrial Organization and Entrepreneurship", in Glaeser, Edward (ed.) *Agglomeration Economics* (Chicago, IL: University of Chicago Press, 2010).
- [71] Rosenthal, Stuart, and William Strange, "The Determinants of Agglomeration", *Journal of Urban Economics* 50 (2001), 191-229.
- [72] Saiz, Albert, "The Geographic Determinants of Housing Supply", *Quarterly Journal of Economics* 125:3 (2010), 1253-1296.
- [73] Samila, Sampsa, and Olav Sorenson, "Venture Capital, Entrepreneurship and Economic Growth", *Review of Economics and Statistics* 93:1 (2011), 338-349.
- [74] Saxenian, AnnaLee, *Regional Advantage: Culture and Competition in Silicon Valley and Route 128* (Cambridge, MA: Harvard University Press, 1994).
- [75] Simon, Curtis, "Human Capital and Metropolitan Employment Growth", *Journal of Urban Economics* 43:2 (1998), 223-243.
- [76] Simon, Curtis, and Clark Nardinelli, "Human Capital and the Rise of American Cities, 1900-1990", *Regional Science and Urban Economics* 32:1 (2002), 59-96.
- [77] Stigler, George, "The Dominant Firm and the Inverted Umbrella", *Journal of Law and Economics* 8 (1965), 167-172.
- [78] Stock, James, and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression", in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, (New York: Cambridge University Press, 2005).
- [79] Storeygard, Adam, "Farther on Down the Road: Transport Costs, Trade and Urban Growth in sub-Saharan Africa", Working Paper (2012).
- [80] van der Ploeg, Frederick, "Natural Resources: Curse or Blessing?", *Journal of Economic Literature* 49:2 (2011), 366-420.
- [81] White, Langdon, "The Iron and Steel Industry of the Pittsburgh District", *Economic Geography* 4:2 (1928), 115-139.
- [82] Whyte, William, *The Organizational Man* (Philadelphia, PA: University of Pennsylvania Press, 1956).
- [83] Wooldridge, Jeffrey, "Score Diagnostics for Linear Models Estimated by Two Stage Least Squares", in *Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao* (Oxford: Blackwell, 1995), 66-87.

## Appendix of IVQR Methodology

Let capital letters denote random variables, and lowercase letters the values these random variables take. We use data for  $n$  observations on a continuous outcome variable  $Y$ , an endogenous variable  $D$ , one or more instruments  $Z$ , and covariates  $X$ . In our setting,  $Y$  is the log employment growth of the city,  $D$  is the initial average establishment size of the city (or the start-up employment share), and  $Z$  are our instruments based upon historical proximity to mines.

We are interested in the potential outcomes  $Y_d$  indexed against the endogenous variable  $d$ . In particular, we look at the conditional quantiles of the potential outcomes,  $\{QY_d(\tau|x), \tau \in (0, 1)\}$ , where  $\tau$  indicates the quantile index and is strictly increasing. The aim is to identify the causal effect of  $D$  on  $Y$  holding unobserved heterogeneity among cities ( $U_D$ ) constant at  $U_D = \tau$ .  $U_D$  is a rank variable that captures the heterogeneity among cities that are similar in terms of their observed characteristics. The causal effect of interest can simply be expressed as  $QY_d(\tau|x) - QY_{d'}(\tau|x)$ . These estimated effects will differ across quantiles  $\tau$  if the effect of the endogenous variable  $D$  on  $Y$  is heterogeneous.

Endogeneity arises as  $D$  and  $U$  are correlated. With endogeneity, the standard quantile regression estimates (Koenker and Bassett, 1978) will be biased and a model with instrumental variables is required. If we have an instrument  $Z$  that is orthogonal to the potential outcome other than through its correlation with the endogeneous variable, we can estimate the causal effect of  $D$  on  $Y$  over the entire distribution of  $Y$ .

Chernozhukov and Hansen (2004a) explain the IVQR model as follows. First, given  $X = x$  the potential outcomes are expressed  $Y = q(d, x, U_d)$ , where the  $\tau^{th}$  quantile  $q(d, x, \tau)$  is strictly increasing and left-continuous in  $\tau$ . Second, given  $X = x$ ,  $\{U_d\}$  is independent of  $Z$ . Third, given  $X = x$  and  $Z = z$ , the potential treatment  $D_z = \delta(z, x, V)$  for any unknown function  $\delta(\cdot)$  and random vector  $V$ . This is the selection equation. Fourth, for each  $d$  and  $d'$ , given  $(V, X, Z)$ ,  $U_d$  is equal in distribution to  $U_{d'}$ . This is the rank similarity assumption, required to interpret the estimates as effects for individual cities at each part of the distribution.<sup>32</sup> Finally, we need to observe  $Y = q(D, X, U_D)$ ,  $D = \delta(Z, X, Y)$ ,  $X$  and  $Z$ .

The model can be estimated using a Matlab procedure as follows. The finite sample quantile regression (QR) objective function is

$$(1) \quad Q_n(\tau, \alpha, \beta, \gamma) := \sum \rho_\tau(Y_c - D'_c\alpha - X'_c\beta - Z'_c\gamma) V_c.$$

Above,  $c$  indexes the observation unit (city),  $D$  is the endogenous variable,  $X$  is the vector of exogenous covariates,  $Z_c = f(X_c, Z_c)$  is the vector of instrumental variables and  $V_c := V(X_c, Z_c) > 0$  is a weight. For a given value ( $\alpha$ ) of the structural parameter, we will run the standard quantile regression to obtain

$$(2) \quad (\hat{\beta}(\alpha, \tau), \hat{\gamma}(\alpha, \tau)) = \arg \min Q_n(\tau, \alpha, \beta, \gamma).$$

---

<sup>32</sup>The rank similarity assumption requires that a city's proximate position in the non-treated distribution is the same (or very similar) to its position in the treated distribution. Achieving rank similarity typically requires conditioning on other important covariates.

To obtain an estimate for  $\alpha(\tau)$ , we find the value of  $\alpha$  that minimizes the absolute value of the instrumental variable,  $\widehat{\gamma}(\alpha, \tau)$ . In other words,

$$(3) \quad \widehat{\alpha}(\tau) = \arg \inf [W_n(\alpha)], \quad W_n(\alpha) := n [\widehat{\gamma}(\alpha, \tau)'] \widehat{A}(\alpha) [\widehat{\gamma}(\alpha, \tau)],$$

where  $\widehat{A} = A(\alpha) + o_p(1)$  and  $A(\alpha)$  is positive definite, uniformly in  $\alpha \in A$ . Setting  $A(\alpha)$  equal to the inverse of the asymptotic covariance matrix of  $\sqrt{n}(\widehat{\gamma}(\alpha\tau) - \gamma(\alpha\tau))$ , the parameter estimates are obtained from

$$(4) \quad \widehat{\theta}(\tau) := \left( \widehat{\alpha}(\tau), \widehat{\beta}(\tau) \right) := \left( \widehat{\alpha}(\tau), \widehat{\beta}(\widehat{\alpha}(\tau), \tau) \right).$$

In short, this is the finite-sample instrumental variable quantile regression (IVQR) using the inverse estimation procedure described in Chernozhukov and Hansen (2004a and 2004b).

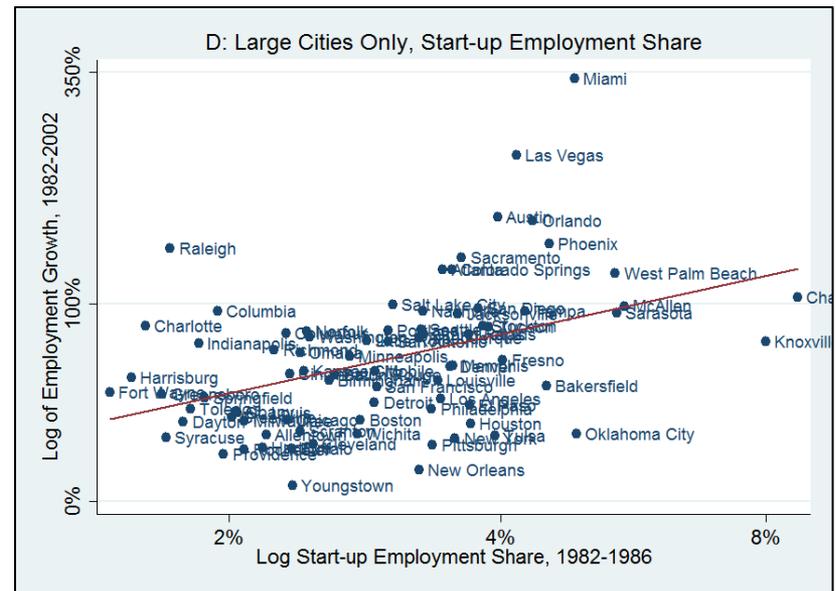
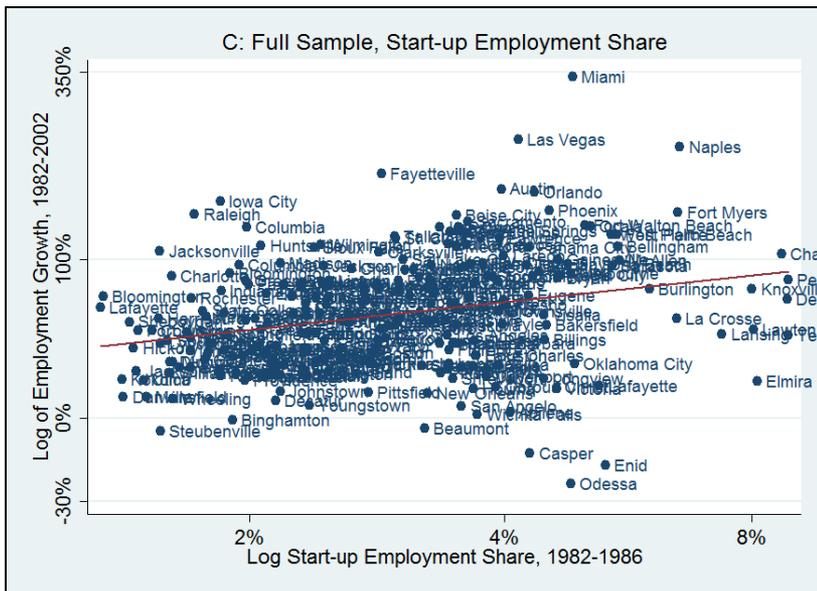
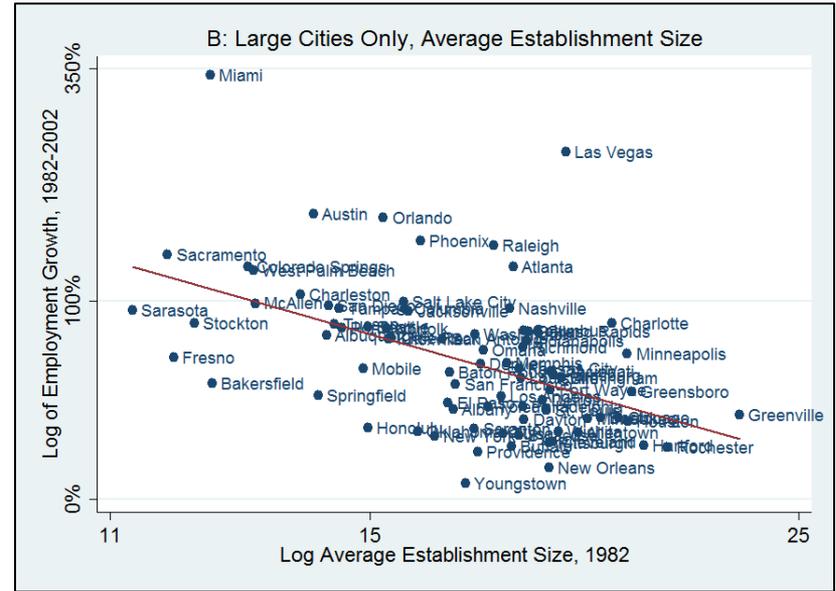
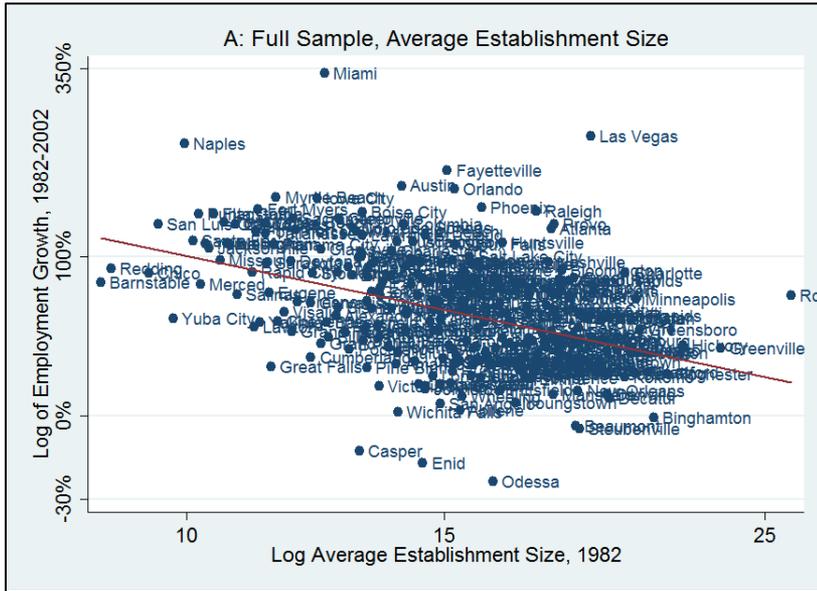
In the context of our city growth estimation,  $Y$  is a measure of employment growth between 1982 and 2002,  $D$  the average establishment size in 1982 (or the start-up employment share in 1982), and  $Z$  are the measures of mine proximity at the beginning of the 20<sup>th</sup> century. The empirical conditional quantile model can be expressed as

$$(5) \quad Qy_c | x(\tau) = \alpha(\tau) + \beta(\tau) D_c + X_c' \lambda(\tau) + \gamma(\tau)_r + \epsilon_c,$$

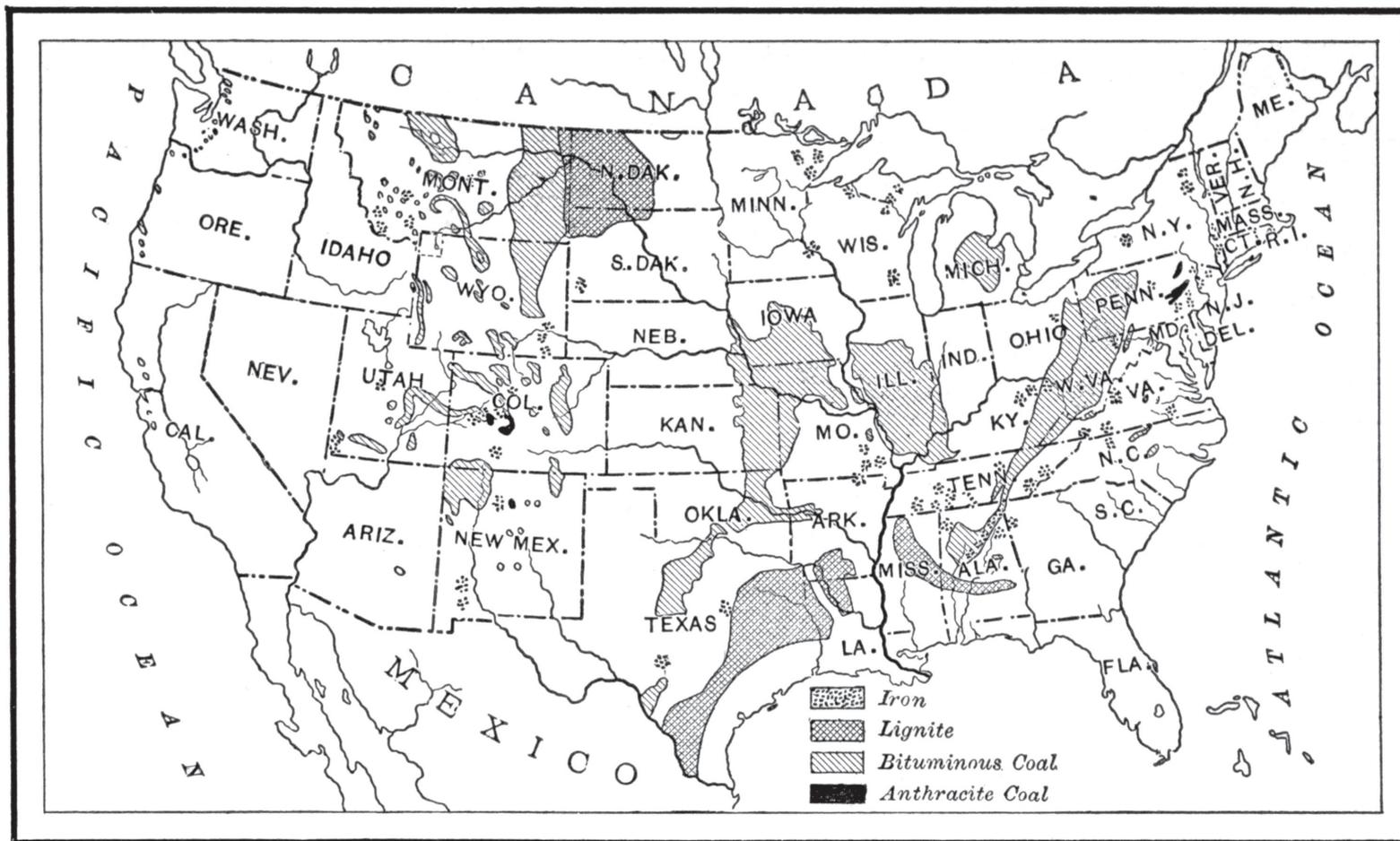
where the subscript  $c$  refers to city and  $r$  to census division. The vector of covariates ( $X_c$ ) includes the initial employment controls and city growth covariates discussed in section 4.

# Fig. 1: City growth and initial entrepreneurship

Cross-sectional plots of urban growth 1982-2002 vs. initial traits



# Fig. 2: Representative map of mineral and coal deposits, 1910



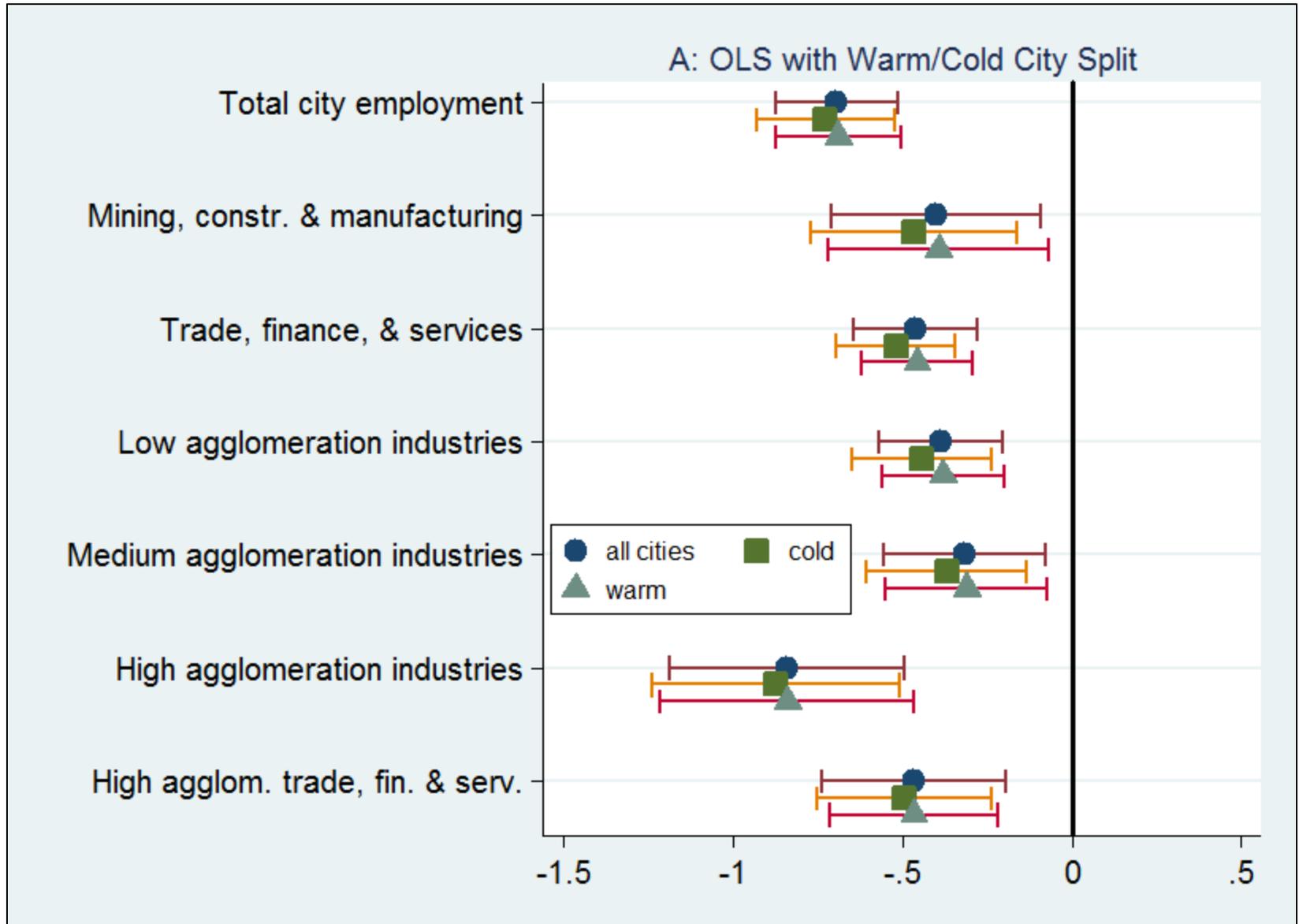
Coal and Iron Deposits in the United States, 1910

Ralph S. Tarr, B.S., F.G.S.A. and Frank M. McMurry, Ph.D., *New Geographies* 2nd ed

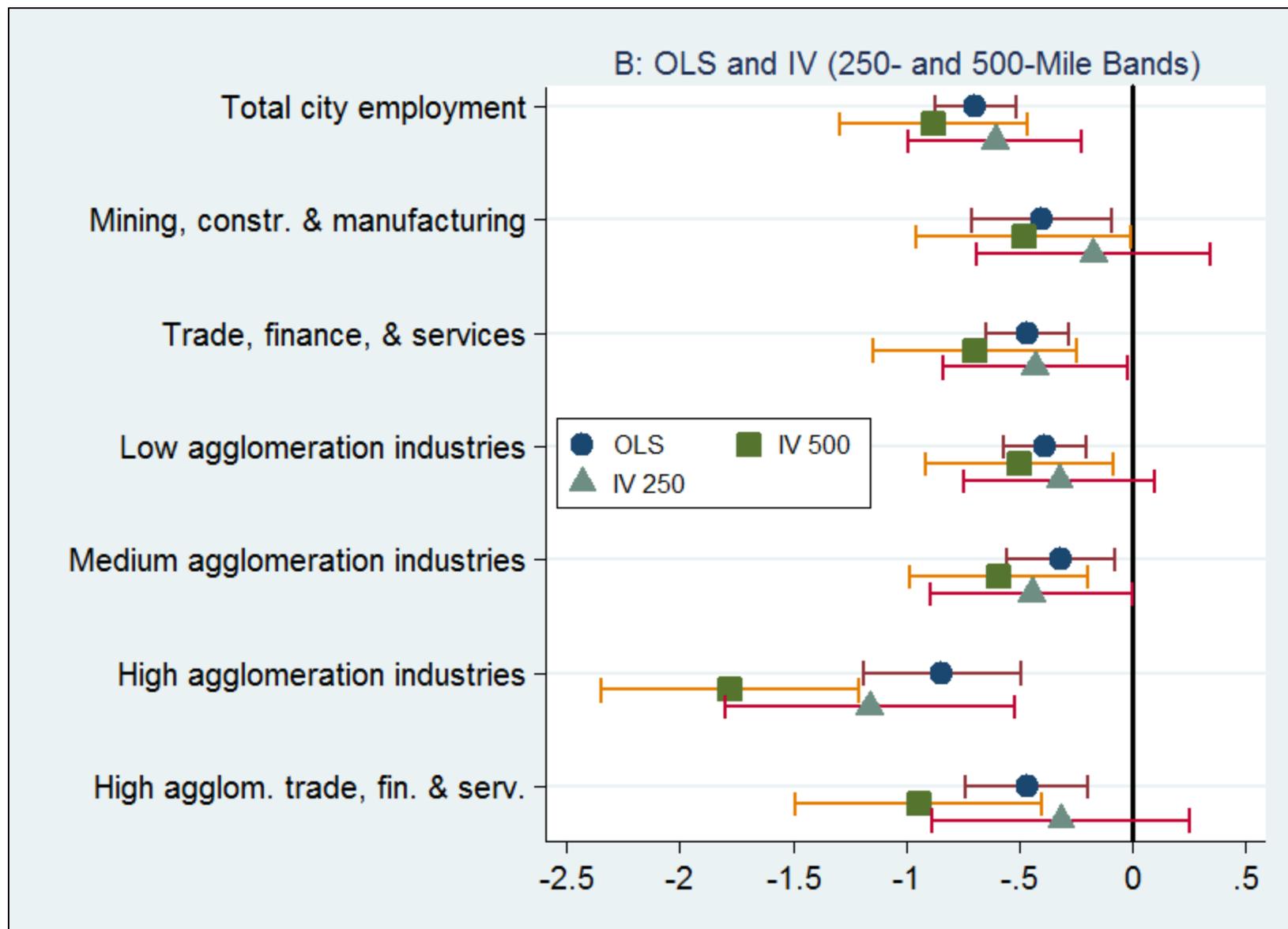
(New York, NY: The Macmillan Company, 1910)

Downloaded from *Maps ETC*, on the web at <http://etc.usf.edu/maps> [map #02085]

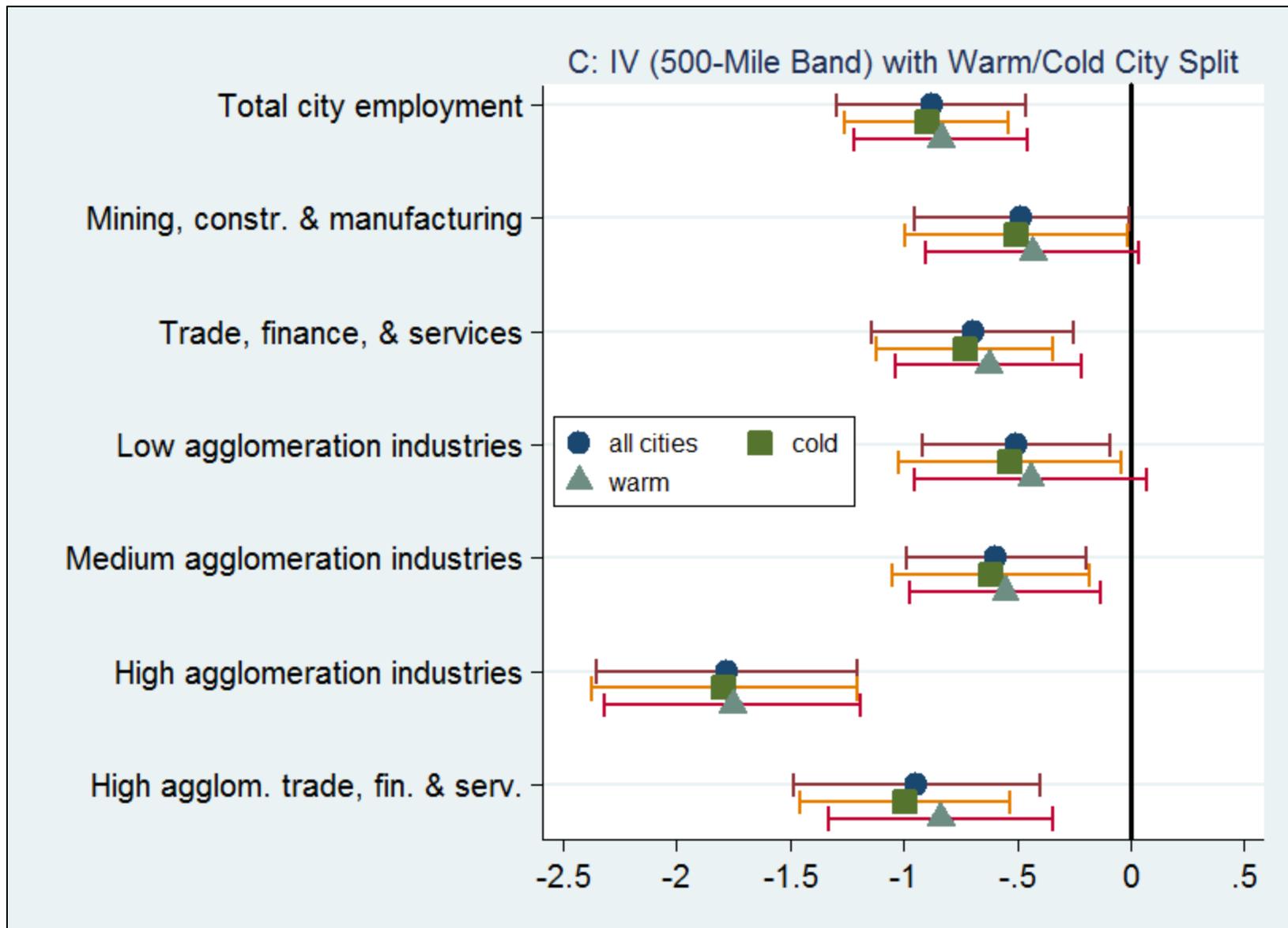
**Fig. 3a: Base estimations of average establishment size effects**  
 Coefficient estimates and 90% confidence intervals



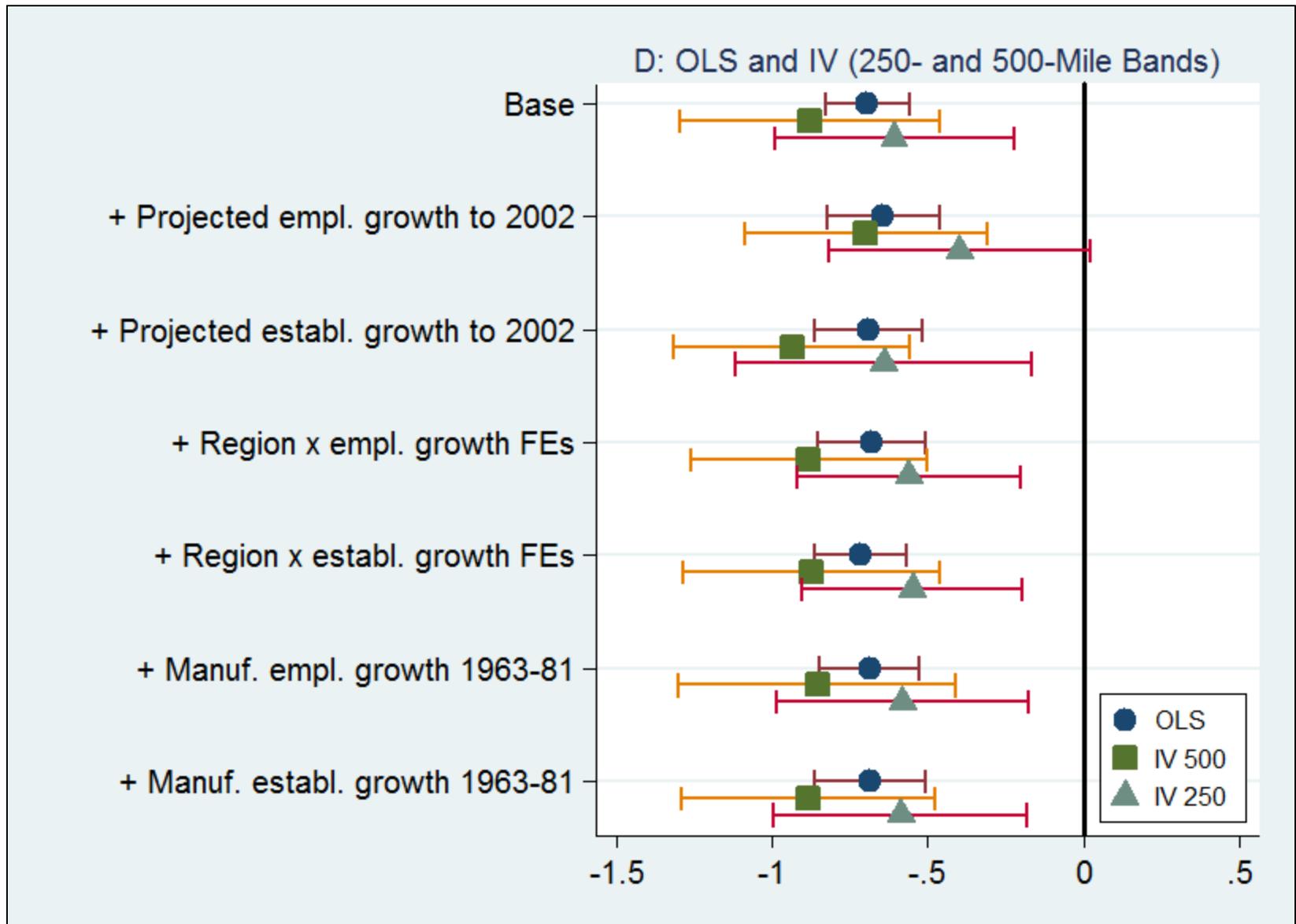
**Fig. 3b: Average establishment size effects for OLS and IV**  
 Coefficient estimates and 90% confidence intervals



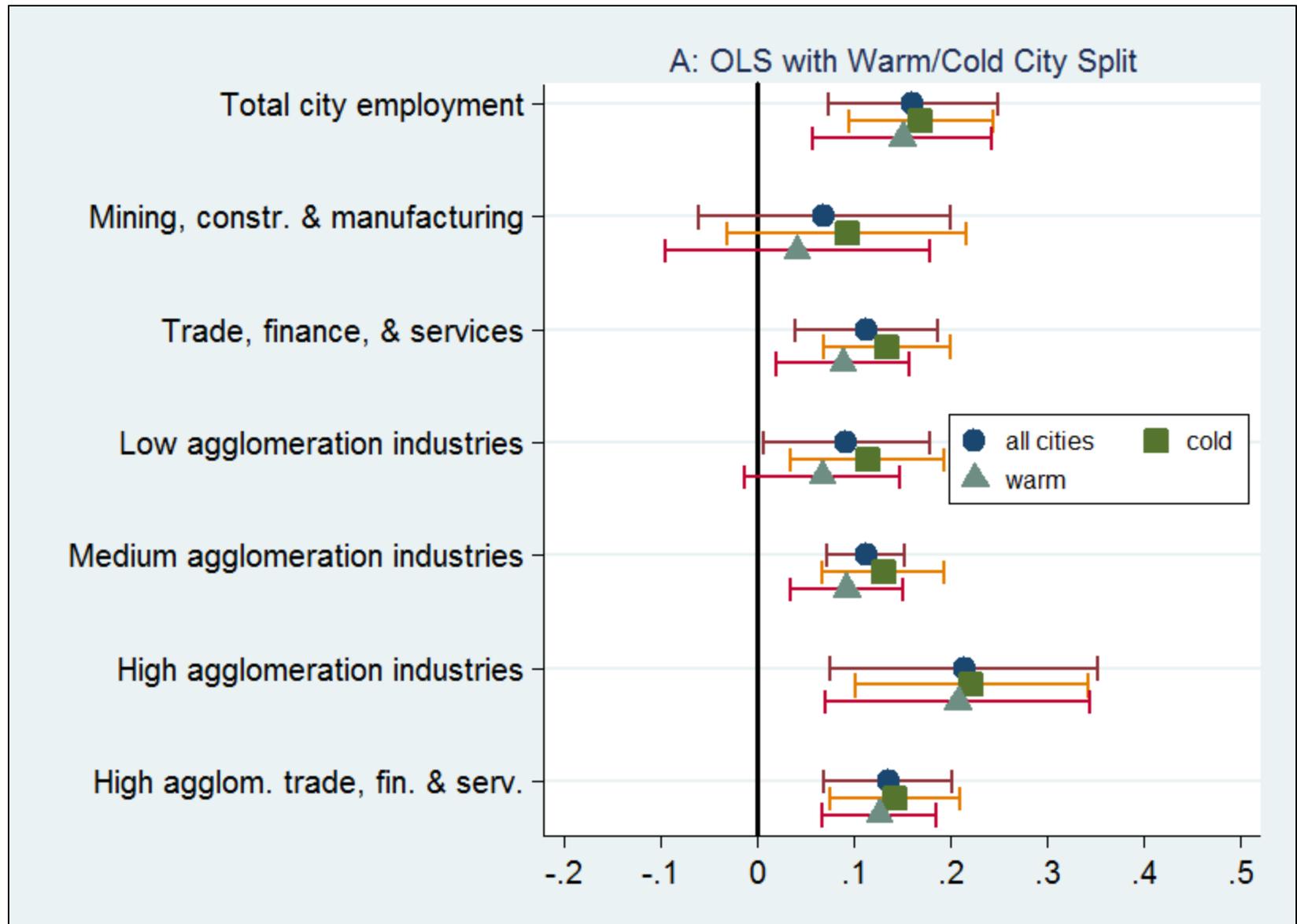
**Fig. 3c: Average establishment size effects with warm/cold IV split**  
 Coefficient estimates and 90% confidence intervals



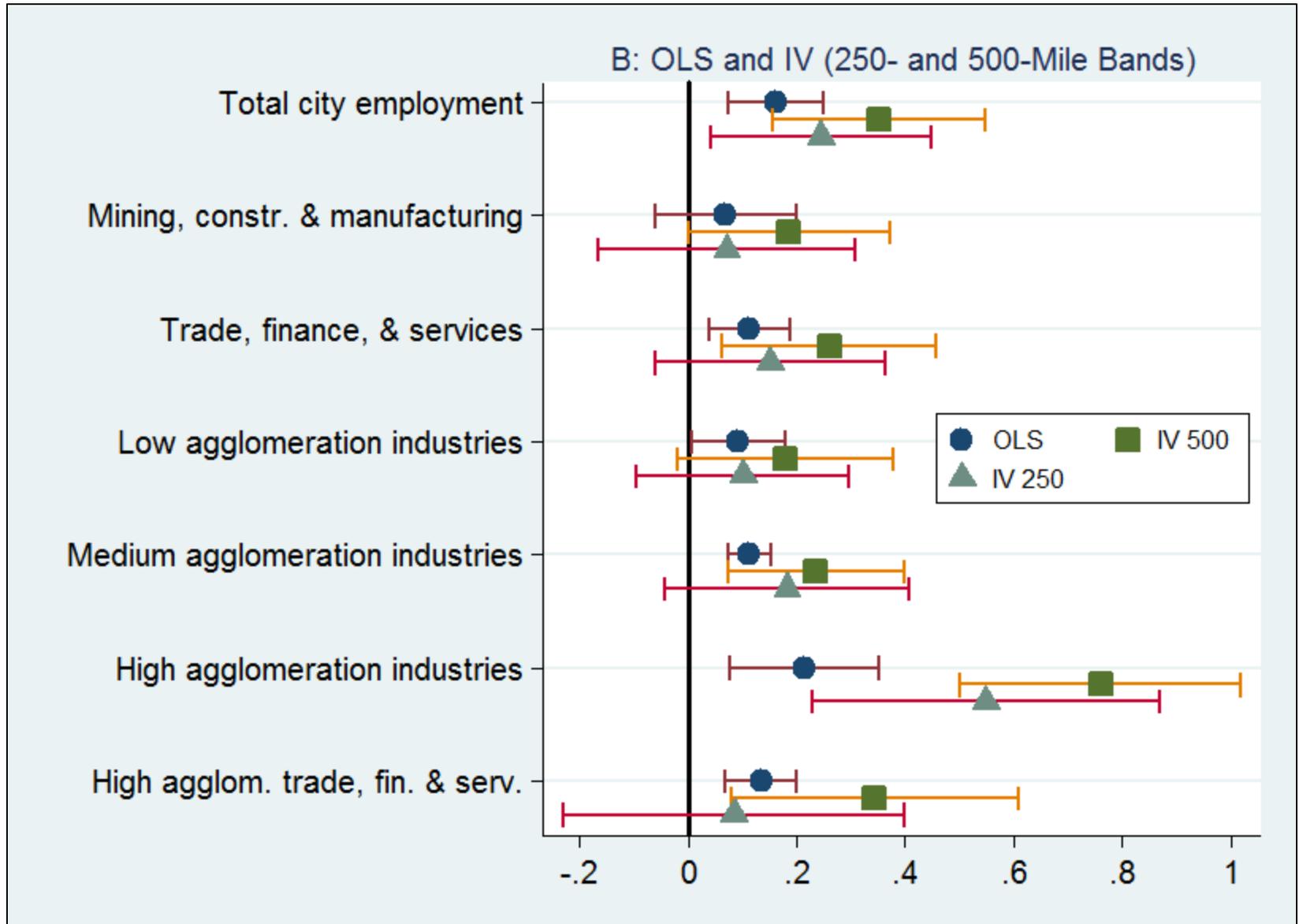
**Fig. 3d: Average establishment effects with extended growth controls**  
 Coefficient estimates and 90% confidence intervals



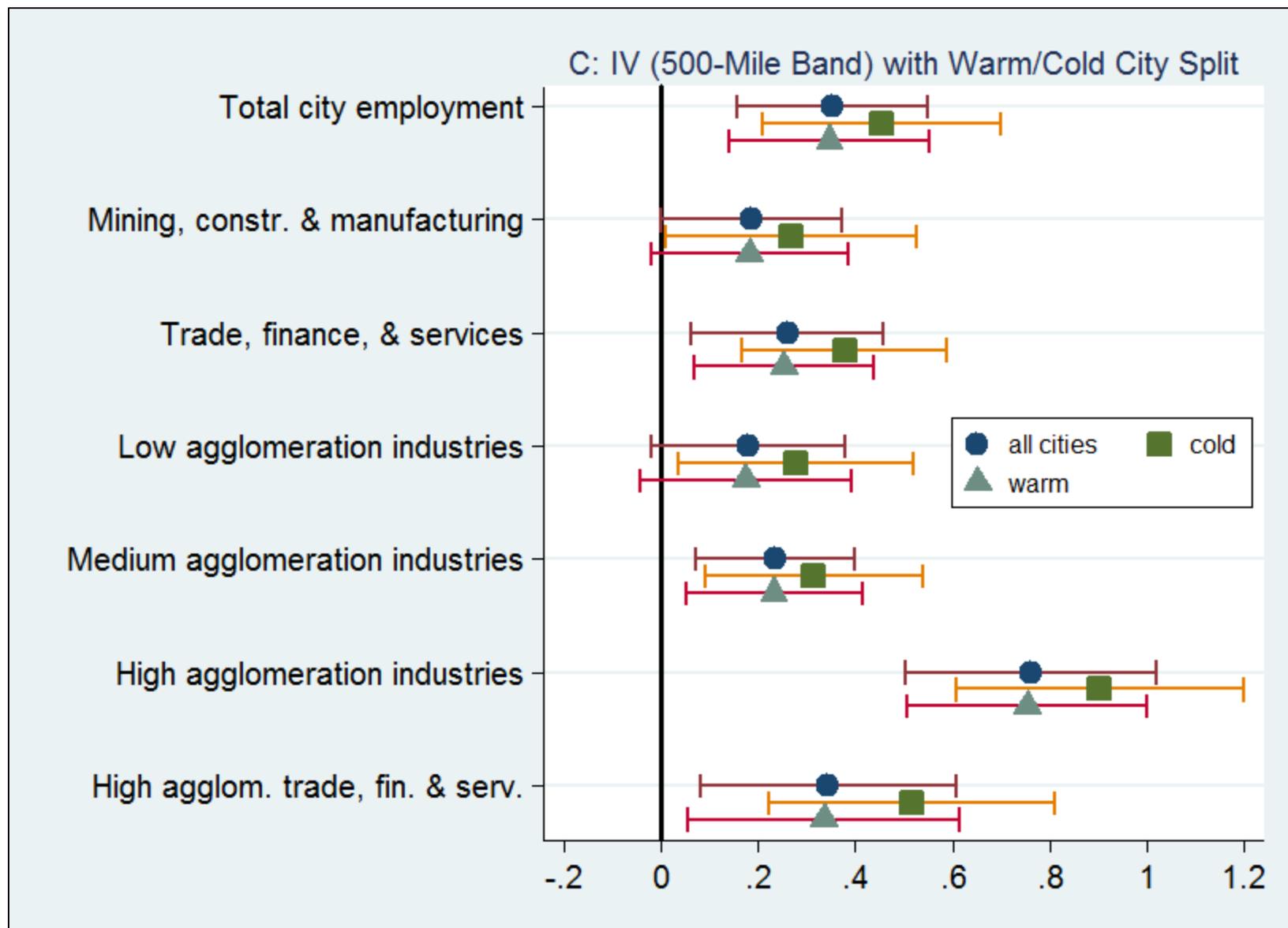
**Fig. 4a: Base estimations of start-up employment share effects**  
 Coefficient estimates and 90% confidence intervals



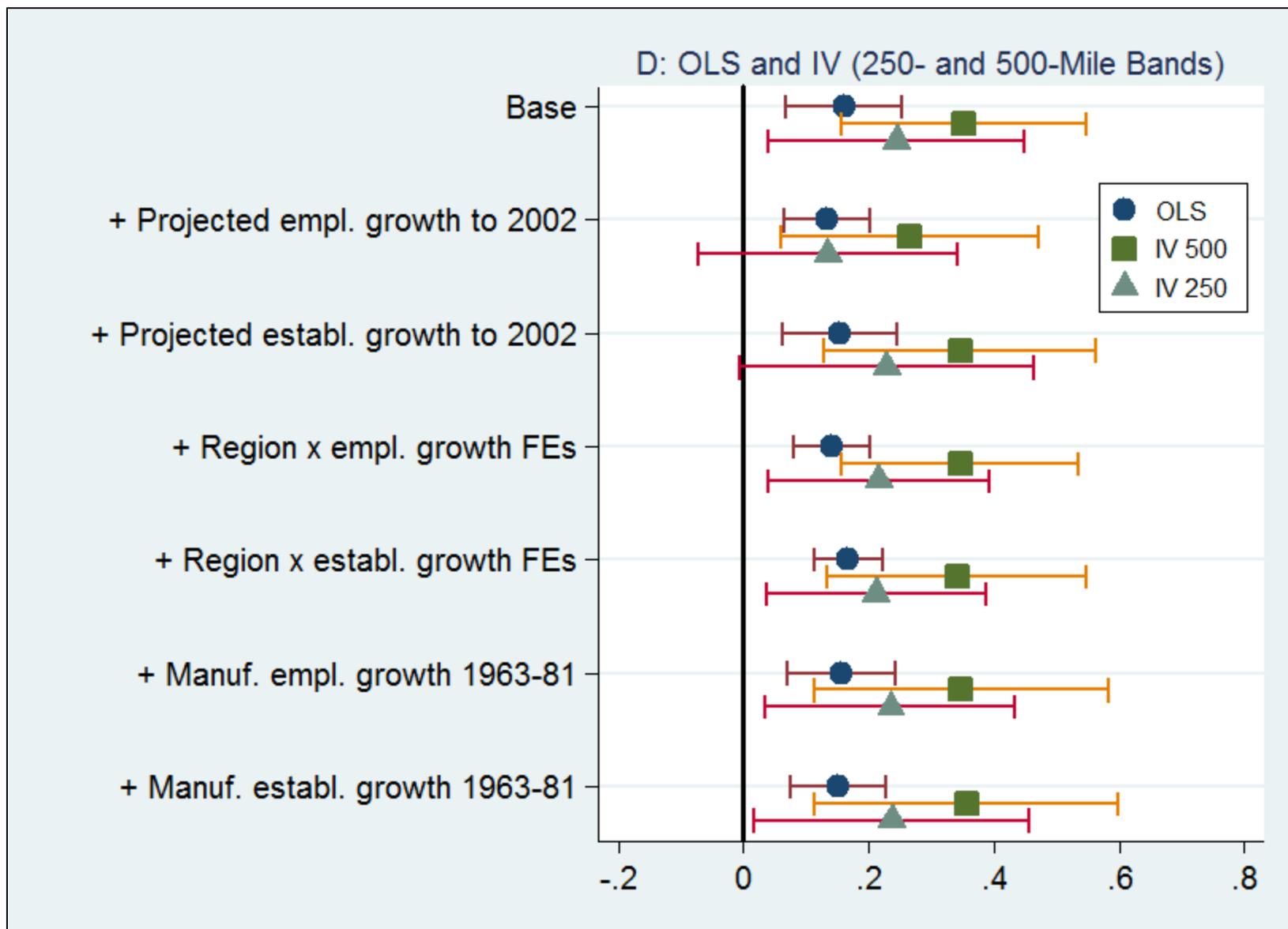
**Fig. 4b: Start-up share effects for OLS and IV**  
 Coefficient estimates and 90% confidence intervals



**Fig. 4c: Start-up share effects with warm/cold IV split**  
 Coefficient estimates and 90% confidence intervals



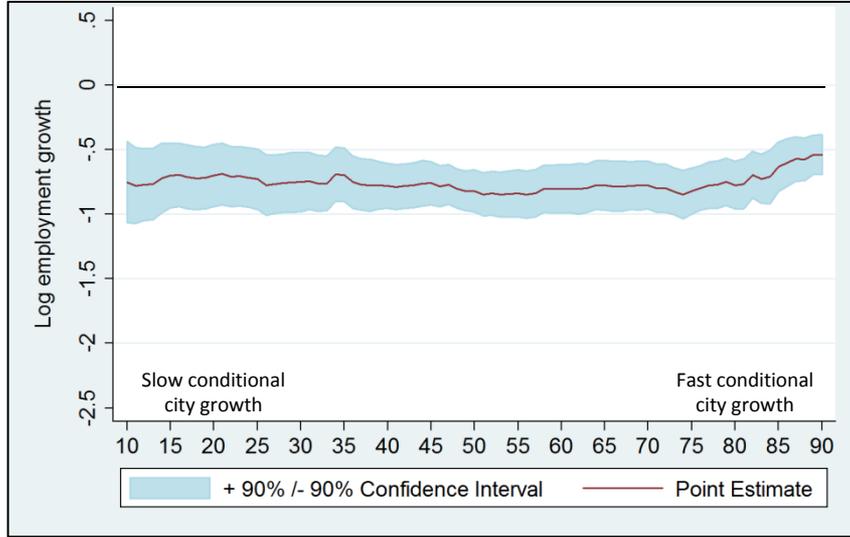
**Fig. 4d: Start-up share effects with extended growth controls**  
 Coefficient estimates and 90% confidence intervals



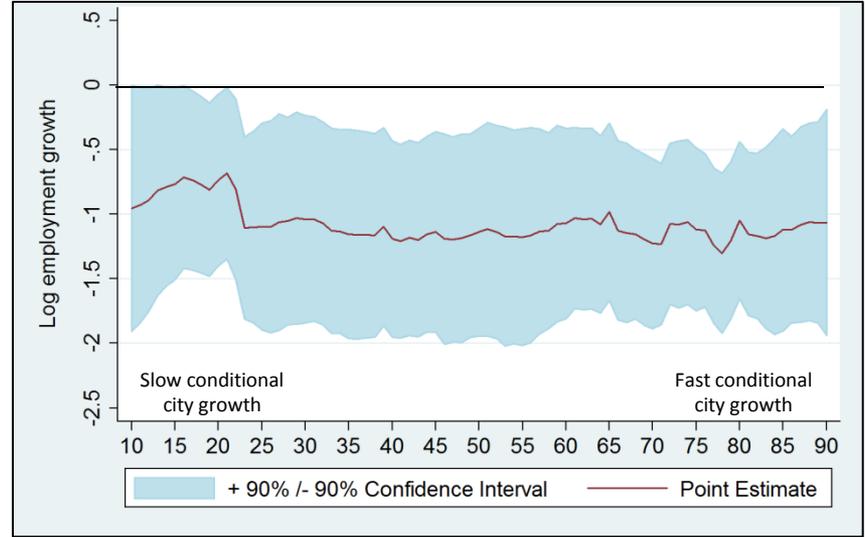
# Fig. 5: Quantile IV estimations of average establishment size effects

## Coefficient estimates and 90% confidence intervals

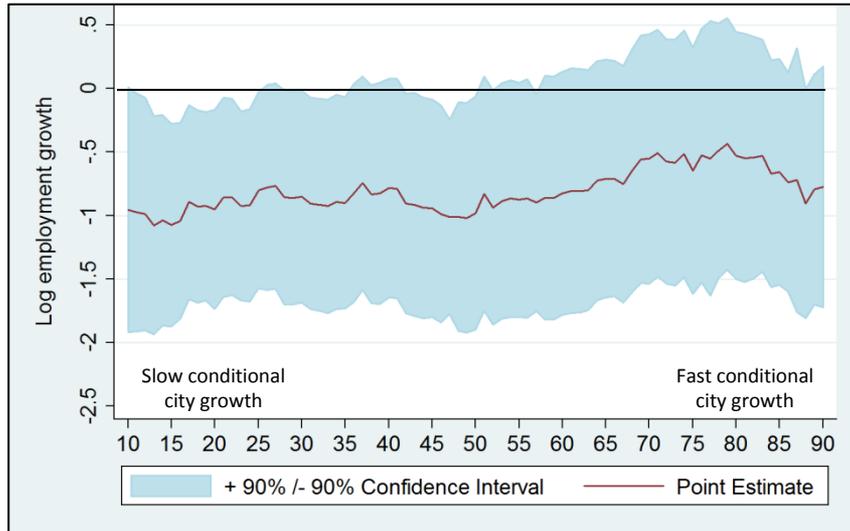
A. Quantile regression estimates



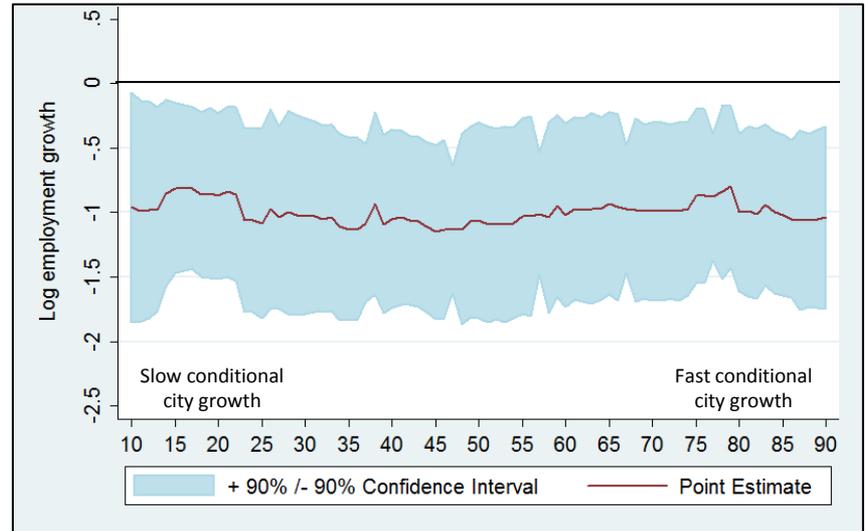
B. IVQR using historical mines and 500-mile band



C. IVQR using historical mines and 250-mile band



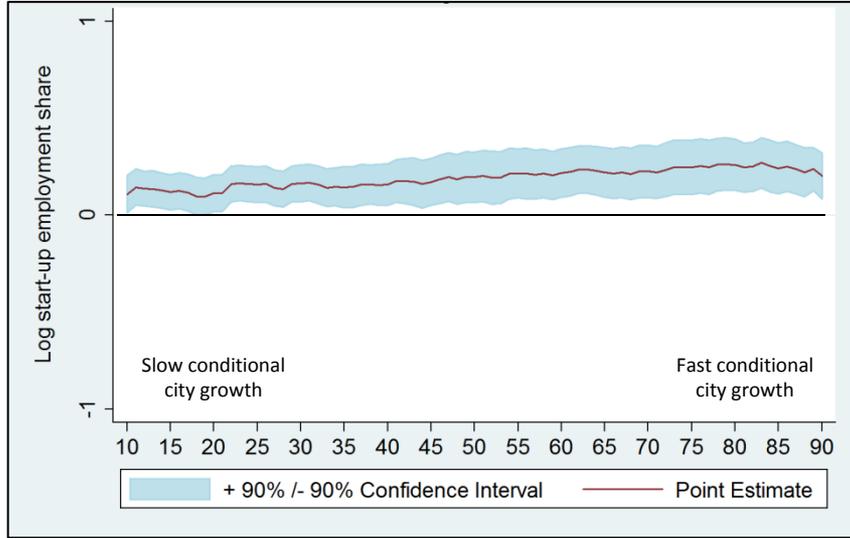
D. IVQR using historical mines and 250- and 500-mile bands



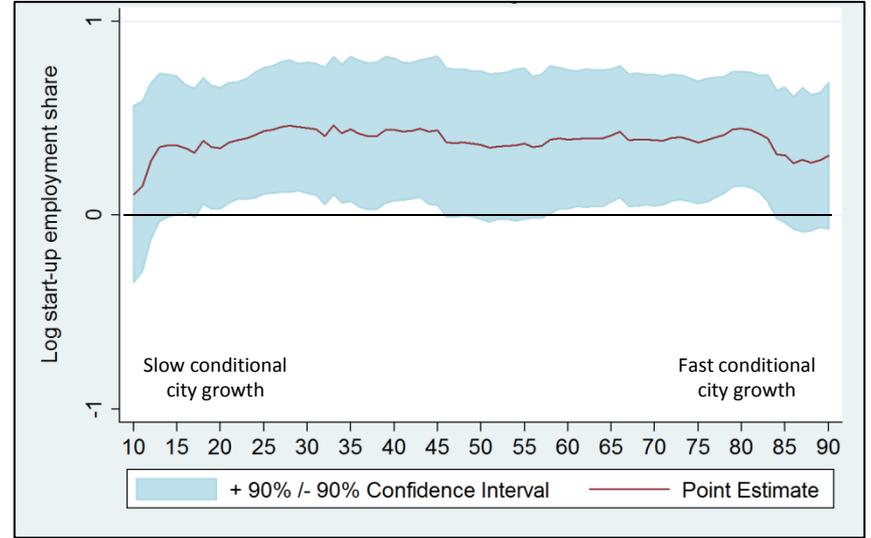
# Fig. 6: Quantile IV estimations of start-up employment share effects

## Coefficient estimates and 90% confidence intervals

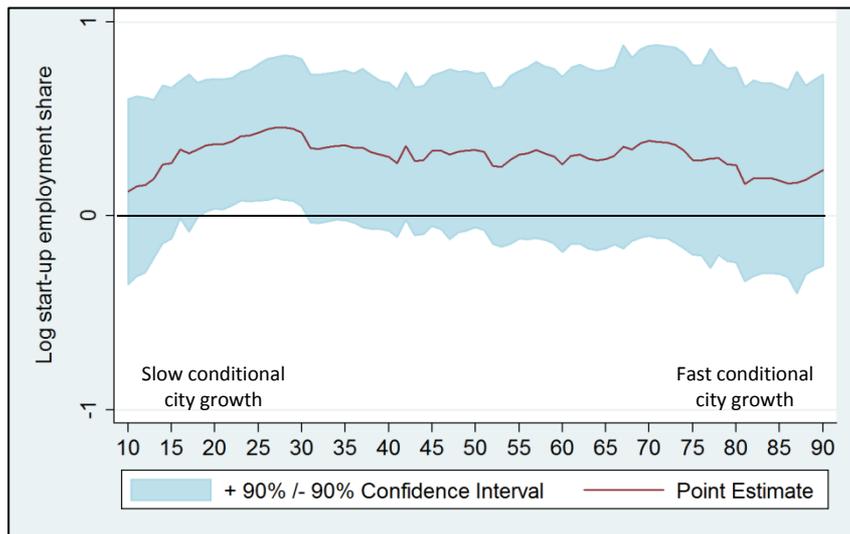
### A. Quantile regression estimates



### B. IVQR using historical mines and 500-mile band



### C. IVQR using historical mines and 250-mile band



### D. IVQR using historical mines and 250- and 500-mile bands

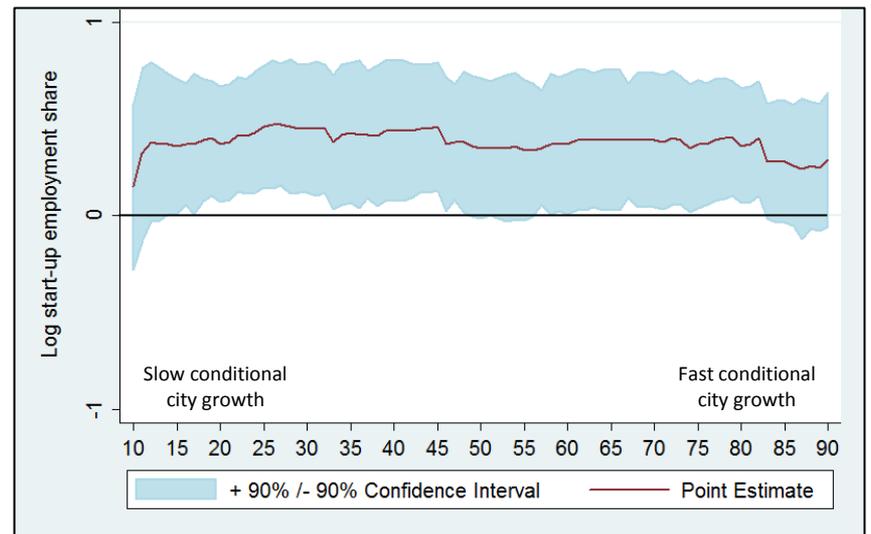


Table 1a: LBD descriptive statistics for cities, circa 1982

	Mean	Standard deviation
City size 1982		
Total employment	231,655	411,379
Mining, construction, & manufacturers	28%	
Trade, finance, & services	72%	
Low agglomeration sectors	57%	
Medium agglomeration sectors	16%	
High agglomeration sectors	26%	
Log employment growth 1982-2002		
Overall	0.361	0.247
Mining, construction, & manufacturers	0.058	0.358
Trade, finance, & services	0.493	0.239
Low agglomeration sectors	0.440	0.233
Medium agglomeration sectors	0.358	0.297
High agglomeration sectors	0.307	0.404
Average establishment size		
Overall, 1982	19.8	3.5
Overall, 2002	19.9	2.8
Mining, construction, & manufacturers, 1982	34.1	14.2
Trade, finance, & services, 1982	15.8	3.2
Start-up share of local firm activity		
Employment, 1982	3.1%	1.6%
Employment, 2002	3.3%	1.3%
Establishment counts, 1982	9.7%	2.2%
Establishment counts, 2002	8.2%	1.7%

Notes: Descriptive statistics from the Longitudinal Business Database for 1982. Jarmin and Miranda (2002) describe the construction of the LBD. Sectors not included are agriculture, forestry and fishing, public administration, the US postal service, and private households. Start-up shares are calculated for the five-year period following the indicated date.

Table 1b: Correlations of entrepreneurship metrics across cities, circa 1982

	Average establishment size			Start-up share of local firm employments	Start-up share of local firm counts
	Overall	Mining, construction, & manufacturers	Trade, finance, & services		
Average establishment size					
Mining, construction, & manufacturers	0.62				
Trade, finance, & services	0.74	0.14			
Start-up share of local firm employments	-0.49	-0.55	-0.15		
Start-up share of local firm counts	-0.41	-0.63	-0.08	0.74	
Share of empl. in small establishments	-0.98	-0.62	-0.71	0.50	0.41

Notes: See Table 1a. Small establishments are defined to be those with 20 or fewer employees. All correlations are significant at a 10% level except the relationship between average size in trade, finance, and services and the start-up share of local firm counts.

Table 2: Entrepreneurship and growth estimations at city level, 1982-2002

	Log employment growth			Log payroll growth			Log wage growth		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Measuring entrepreneurship through average establishment size									
Log average establishment size in city at start of period	-0.566 (0.078)	-0.598 (0.072)	-0.693 (0.082)	-0.435 (0.092)	-0.478 (0.120)	-0.640 (0.122)	0.073 (0.041)	0.018 (0.063)	-0.054 (0.036)
Initial employment controls		Yes	Yes		Yes	Yes		Yes	Yes
Census division fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
City growth covariates			Yes			Yes			Yes
B. Measuring entrepreneurship through start-up employment share									
Log start-up share of employment in city at start of period	0.200 (0.037)	0.200 (0.053)	0.161 (0.056)	0.200 (0.046)	0.196 (0.060)	0.150 (0.065)	0.016 (0.017)	0.029 (0.016)	0.019 (0.020)
Initial employment controls		Yes	Yes		Yes	Yes		Yes	Yes
Census division fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
City growth covariates			Yes			Yes			Yes

Notes: Estimations describe the OLS relationship between entrepreneurship and city growth. City growth is calculated as the log ratio of employments at the end of the period to the beginning of the period. Regressions are unweighted, report bootstrapped standard errors, and have 291 observations. Initial employment controls are log employment levels at start of period and their squared values. City growth covariates include log January temperature, log July temperature, log 1970 share of workers with bachelor's education or higher, log 1970 population density, log 1970 population, and log 1970 housing prices. Nine census divisions are used in the fixed effects. A 2% winsorization is employed on variables. Appendix Table 1 repeats these estimations by time period.

Table 3: Sample decomposition of city level estimations for log employment growth

	Total employment growth	Mining, construction & mfg. sectors	Trade, finance, & services sectors	Level of industry agglomeration			High agglom. trade, finance, & services
				Low	Medium	High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Measuring entrepreneurship through average establishment size							
Log average establishment size in city at start of period	-0.693 (0.110)	-0.400 (0.187)	-0.462 (0.110)	-0.387 (0.110)	-0.316 (0.144)	-0.842 (0.210)	-0.467 (0.164)
B. Panel A with warm/cold city interaction							
Log average establishment size in city at start of period	-0.726 (0.124)	-0.466 (0.184)	-0.519 (0.107)	-0.441 (0.125)	-0.369 (0.144)	-0.874 (0.221)	-0.492 (0.156)
x indicator variable for warm city (January temperature >34)	0.037 (0.019)	0.073 (0.079)	0.063 (0.015)	0.061 (0.020)	0.059 (0.020)	0.035 (0.064)	0.027 (0.019)
Linear combination for warm cities	-0.689 (0.112)	-0.393 (0.197)	-0.456 (0.100)	-0.380 (0.110)	-0.310 (0.145)	-0.839 (0.227)	-0.465 (0.150)
C. Measuring entrepreneurship through start-up employment share							
Log start-up share of employment in city at start of period	0.161 (0.053)	0.069 (0.079)	0.113 (0.045)	0.092 (0.052)	0.112 (0.024)	0.214 (0.084)	0.135 (0.040)
D. Panel C with warm/cold city interaction							
Log start-up share of employment in city at start of period	0.169 (0.045)	0.093 (0.075)	0.134 (0.040)	0.114 (0.048)	0.130 (0.038)	0.221 (0.073)	0.142 (0.041)
x indicator variable for warm city (January temperature >34)	-0.019 (0.030)	-0.053 (0.022)	-0.045 (0.019)	-0.047 (0.015)	-0.038 (0.018)	-0.013 (0.023)	-0.016 (0.016)
Linear combination for warm cities	0.150 (0.056)	0.041 (0.083)	0.088 (0.042)	0.067 (0.049)	0.092 (0.035)	0.207 (0.083)	0.126 (0.036)

Notes: See Table 2. Regressions include Initial employment controls, Census division fixed effects, and City growth covariates. Warm cities are defined to be the 143 cities with January temperatures above the median for the United States of 34 degrees.

Table 4: Entrepreneurship and growth regressions at city-industry level, 1982-2002

	Total employment	Total employment	Mining, construction, & mfg.	Trade, finance, & services	Cold cities Jan. temp <34	Warm cities Jan. temp >34	1982-1992 period	1992-2002 period
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Measuring entrepreneurship through average establishment size								
Log average establishment size in city-industry at start of period	-0.192 (0.015)	-0.165 (0.015)	-0.291 (0.015)	-0.120 (0.017)	-0.158 (0.022)	-0.175 (0.018)	-0.104 (0.008)	-0.095 (0.008)
Initial employment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region x industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City growth covariates	Yes							
City fixed effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes
B. Measuring entrepreneurship through start-up employment share								
Log start-up share of employment in city-industry at start of period	0.054 (0.007)	0.042 (0.007)	0.055 (0.011)	0.039 (0.006)	0.036 (0.006)	0.049 (0.010)	0.027 (0.004)	0.019 (0.004)
Initial employment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region x industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City growth covariates	Yes							
City fixed effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: See Table 2. Estimations describe the OLS relationship between entrepreneurship and city-industry growth. Industries are defined at the two-digit level of the SIC system. Region x industry fixed effects use the nine census divisions. Initial employment controls are city-industry specific. City-industries must have 100 employees throughout the 1977-2002 period to be included in the full sample, for 12,178 observations.

Table 5a: Coal prices and distance from mines in 1925-1930, anthracite

	Indicator variable for anthracite price listed		Log price of anthracite coal in city	
	(1)	(2)	(3)	(4)
Count of anthracite mines 0-50 miles	0.067 (0.026)	0.051 (0.011)	-0.043 (0.017)	-0.052 (0.009)
Count of anthracite mines 50-100 miles	0.053 (0.015)	0.062 (0.018)	-0.021 (0.010)	-0.018 (0.011)
Count of anthracite mines 100-250 miles	0.163 (0.027)	0.186 (0.031)	-0.032 (0.016)	-0.041 (0.013)
Count of anthracite mines 250-500 miles	0.062 (0.060)	0.089 (0.074)	-0.009 (0.021)	-0.025 (0.006)
Year fixed effects	Yes	Yes	Yes	Yes
Census division fixed effects		Yes		Yes
Observations	261	261	133	133

Notes: Estimations describe the OLS relationship between anthracite coal prices observed in cities in 1925-1930 and their distances from mines. Included city-year observations list an anthracite price, a bituminous price, or both. Price data are available for 47 cities in our sample. Columns 1 and 2 present linear probability models that an anthracite price is listed. Columns 3 and 4 consider the log price of anthracite coal when listed. Explanatory variables are anthracite mine counts by spatial bands from cities; counts are transformed to have unit standard deviation for interpretation. Regressions are unweighted and cluster standard errors by city.

Table 5b: Coal prices and distance from mines in 1925-1930, bituminous

	Indicator variable for bituminous price listed		Log price of bituminous coal in city	
	(1)	(2)	(3)	(4)
Count of bituminous mines 0-50 miles	0.027 (0.013)	0.006 (0.017)	-0.043 (0.012)	-0.036 (0.013)
Count of bituminous mines 50-100 miles	0.026 (0.023)	0.134 (0.040)	-0.053 (0.020)	0.012 (0.032)
Count of bituminous mines 100-250 miles	0.044 (0.039)	-0.019 (0.028)	-0.121 (0.027)	-0.128 (0.035)
Count of bituminous mines 250-500 miles	0.026 (0.042)	0.049 (0.039)	-0.142 (0.049)	-0.124 (0.035)
Year fixed effects	Yes	Yes	Yes	Yes
Census division fixed effects		Yes		Yes
Observations	261	261	216	216

Notes: See Table 5a. The analysis is repeated for bituminous coal.

Table 6: Historical mining deposits and the development of industrial structures

	Log average establishment size in manuf. 1963	Log average employments in mining near city 1976-1980	Log average establishment size in city, 1982			Log start-up employment share in city 1982
			Total	Mining, constr., & manufacturing sectors	Trade, finance, & services sectors	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Log mine counts within 500 miles of city, 1900						
Log mine counts within 500 miles of city, 1900	0.213 (0.045)	1.163 (0.054)	0.075 (0.011)	0.142 (0.022)	0.042 (0.010)	-0.161 (0.021)
B. Panel A including indicator variable for coal or iron ore being top mineral product of state, 1928						
Log mine counts within 500 miles of city, 1900	0.204 (0.039)	1.149 (0.049)	0.071 (0.009)	0.133 (0.022)	0.037 (0.011)	-0.154 (0.023)
(0,1) Coal or iron ore is the top mineral product of the state, 1928	0.126 (0.069)	0.187 (0.058)	0.062 (0.017)	0.116 (0.041)	0.060 (0.018)	-0.096 (0.044)
C. Log 1 + mine counts within 250 miles of city, 1900						
Log 1 + mine counts within 250 miles of city, 1900	0.087 (0.020)	0.476 (0.047)	0.034 (0.006)	0.066 (0.014)	0.010 (0.007)	-0.073 (0.014)
D. Count of mines within distance rings of city, 1900 (coefficients x100)						
Mine counts within 100 miles of city, 1900	0.085 (0.027)	0.328 (0.060)	0.026 (0.007)	0.065 (0.019)	0.000 (0.008)	-0.091 (0.019)
Mine counts within 100-500 miles of city, 1900	0.039 (0.008)	0.198 (0.015)	0.011 (0.002)	0.020 (0.006)	0.009 (0.002)	-0.030 (0.006)
E. Indicator variable for coal composition within 150 miles that favors anthracite or bituminous, present						
(0,1) Composition of coal within 150 miles favors anthracite or bituminous	0.378 (0.069)	1.143 (0.145)	0.095 (0.018)	0.234 (0.047)	0.031 (0.020)	-0.207 (0.049)

Notes: See Table 2. Regressions include Initial employment controls, Census division fixed effects, and City growth covariates. Appendix Tables 2a and 2b provide descriptive statistics regarding mining counts. Coal or iron ore is the top mineral product in 1928 for AL, CO, IL, IN, KY, MD, MI, MN, ND, PA, TN, VA, WA, and WV.

Table 7: City level IV estimations of employment growth

	Instruments employed in estimation:						
	Yes	Yes	Yes		Yes	Yes	
Log count of mines 500 miles, 1900	Yes	Yes	Yes				
(0,1) coal or iron ore top product, 1928		Yes	Yes		Yes	Yes	
Count of iron miles 100 miles, 1900			Yes		Yes	Yes	
Log 1+count of mines 250 miles, 1900				Yes	Yes	Yes	
Log count of mines 500 miles, present							Yes
Coal composition 150 miles, present							Yes
Count of iron miles 100 miles, present							Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Measuring entrepreneurship through average establishment size							
Log average establishment size in city at start of period	-0.967 (0.311)	-0.930 (0.280)	-0.878 (0.251)	-0.517 (0.325)	-0.608 (0.232)	-0.783 (0.241)	-0.831 (0.258)
First stage partial R squared	0.151	0.186	0.193	0.106	0.175	0.208	0.162
F test statistic in first stage	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum 2SLS relative bias	<10%	<10%	<10%	<10%	<15%	<15%	<15%
Over-identification test p-value		0.689	0.357		0.624	0.124	0.457
Exogeneity test p-value	0.168	0.177	0.281	0.470	0.642	0.585	0.471
B. Measuring entrepreneurship through start-up employment share							
Log start-up share of employment in city at start of period	0.450 (0.144)	0.455 (0.156)	0.352 (0.119)	0.245 (0.155)	0.245 (0.124)	0.335 (0.125)	0.362 (0.137)
First stage partial R squared	0.114	0.128	0.157	0.078	0.129	0.154	0.129
F test statistic in first stage	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum 2SLS relative bias	<10%	<15%	<15%	<10%	<15%	<20%	<15%
Over-identification test p-value		0.901	0.083		0.083	0.069	0.341
Exogeneity test p-value	0.004	0.002	0.024	0.440	0.376	0.043	0.034

Notes: See Tables 2 and 6. Outcome variable is log employment growth for cities. Instruments are indicated by column titles. Regressions are unweighted, report bootstrapped standard errors, and have 291 observations. Regressions include Initial employment controls, Census division fixed effects, and City growth covariates. The maximum 2SLS relative bias reports the minimum bias that can specified and still reject the null hypothesis that the instruments are weak. This level is determined through the minimum eigenvalue statistic and Stock and Yogo's (2005) 2SLS size of nominal 5% Wald test. The null hypothesis in Basman's over-identification tests is that the instruments are valid. The null hypothesis in Wu-Hausman exogeneity tests is that the instrumented regressors are exogenous.

Table 8a: Extended city outcomes using average establishment size

	Log employment in 2002 from entrants since 1982 relative to city size in 1982	Employment share in 2002 of entrants since 1982 by size category in 2002			Log average age of establishments in city in 2002	
		All	>100 employees	<=100 employees	All	Top 25 employers
	(1)	(2)	(3)	(4)	(5)	(6)
A. OLS relationship using average establishment size						
Log average establishment size in city at start of period	-0.937 (0.134)	-0.139 (0.030)	-0.133 (0.044)	-0.051 (0.018)	0.202 (0.046)	0.140 (0.077)
B. IV relationship with triple instrument using 500-mile band						
Log average establishment size in city at start of period	-1.208 (0.293)	-0.216 (0.059)	-0.265 (0.090)	-0.133 (0.042)	0.365 (0.085)	0.460 (0.130)
First stage statistics	Partial R squared 0.193; F statistic 0.000; Maximum 2SLS relative bias <15%					
Over-identification test p-value	0.596	0.938	0.878	0.213	0.943	0.194
Exogeneity test p-value	0.188	0.080	0.067	0.007	0.011	0.003
C. IV relationship with triple instrument using 250-mile band						
Log average establishment size in city at start of period	-0.951 (0.284)	-0.206 (0.061)	-0.292 (0.086)	-0.053 (0.034)	0.344 (0.075)	0.536 (0.153)
First stage statistics	Partial R squared 0.175; F statistic 0.000; Maximum 2SLS relative bias <15%					
Over-identification test p-value	0.756	0.892	0.933	0.895	0.856	0.394
Exogeneity test p-value	0.951	0.147	0.038	0.942	0.036	0.001

Notes: See Tables 2 and 7. Outcome variables are indicated by column headers. Column 1 reports the log net employment growth by entrants in 2002 compared to 1982 city size. Columns 2-4 consider employment shares in 2002 of entrants since 1982 by establishment size category in 2002. Columns 5 and 6 report average establishment ages in 2002 for cities with a maximum age of 26 years. The instruments are the log count of mines within 250 or 500 miles, 1900, as indicated; a (0,1) indicator variable for coal or iron ore being the top mineral product in the state, 1928; and the count of iron ore mines within 100 miles of the city, 1900. Regressions include Initial employment controls, Census division fixed effects, and City growth covariates.

Table 8b: Extended city outcomes using start-up employment share

	Log employment in 2002 from entrants since 1982 relative to city size in 1982	Employment share in 2002 of entrants since 1982 by size category in 2002			Log average age of establishments in city in 2002	
		All	>100 employees	<=100 employees	All	Top 25 employers
		(1)	(2)	(3)	(4)	(5)
A. OLS relationship using start-up employment share						
Log start-up share of employment in city at start of period	0.216 (0.085)	0.034 (0.016)	0.029 (0.023)	0.025 (0.007)	-0.062 (0.019)	-0.059 (0.021)
B. IV relationship with triple instrument using 500-mile band						
Log start-up share of employment in city at start of period	0.497 (0.150)	0.095 (0.027)	0.111 (0.038)	0.058 (0.018)	-0.160 (0.051)	-0.194 (0.064)
First stage statistics	Partial R squared 0.157; F statistic 0.000; Maximum 2SLS relative bias <15%					
Over-identification test p-value	0.140	0.720	0.502	0.139	0.593	0.093
Exogeneity test p-value	0.007	0.003	0.014	0.018	0.001	0.007
C. IV relationship with triple instrument using 250-mile band						
Log start-up share of employment in city at start of period	0.403 (0.140)	0.095 (0.028)	0.126 (0.048)	0.026 (0.018)	-0.158 (0.043)	-0.237 (0.074)
First stage statistics	Partial R squared 0.175; F statistic 0.000; Maximum 2SLS relative bias <15%					
Over-identification test p-value	0.251	0.712	0.440	0.997	0.593	0.175
Exogeneity test p-value	0.109	0.009	0.009	0.939	0.004	0.001

Notes: See Table 8a.

App. Table 1: Basic entrepreneurship and city employment growth regressions by time period

	1982-2002			1982-1992			1992-2002		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Measuring entrepreneurship through average establishment size									
Log average establishment size in city at start of period	-0.566 (0.078)	-0.598 (0.072)	-0.693 (0.082)	-0.385 (0.056)	-0.407 (0.075)	-0.432 (0.113)	-0.216 (0.049)	-0.280 (0.057)	-0.339 (0.044)
Initial employment controls		Yes	Yes		Yes	Yes		Yes	Yes
Census division fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
City growth covariates			Yes			Yes			Yes
B. Measuring entrepreneurship through start-up employment share									
Log start-up share of employment in city at start of period	0.200 (0.037)	0.200 (0.053)	0.161 (0.056)	0.099 (0.030)	0.128 (0.047)	0.110 (0.038)	0.123 (0.023)	0.087 (0.048)	0.065 (0.050)
Initial employment controls		Yes	Yes		Yes	Yes		Yes	Yes
Census division fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
City growth covariates			Yes			Yes			Yes

Notes: See Table 2.

App. Table 2a: Descriptive statistics on known mineral mines, 1900

	Mean	Minimum	Maximum	Maximum City
Total mine counts recorded as discovered by 1900				
Total 0-100 miles	56	0	683	Hickory-Morganton-Lenoir, NC
Total 0-250 miles	270	0	1,022	Greensboro-Winston-Salem-High Point, NC
Total 0-500 miles	946	10	2,966	Salt Lake City-Ogden, UT
100-250 miles	214	0	990	Charleston, WV
250-500 miles	676	6	2,282	Provo-Orem, UT
Mines related to iron ore recorded as discovered by 1900				
Total 0-100 miles	13	0	232	Hickory-Morganton-Lenoir, NC
Total 0-250 miles	67	0	321	Roanoke, VA
Total 0-500 miles	244	2	630	Cleveland-Akron, OH
100-250 miles	54	0	314	Charleston, WV
250-500 miles	177	0	621	Toledo, OH
Mines not related to iron ore recorded as discovered by 1900				
Total 0-100 miles	43	0	451	Hickory-Morganton-Lenoir, NC
Total 0-250 miles	203	0	837	Reno, NV
Total 0-500 miles	701	6	2,629	Salt Lake City-Ogden, UT
100-250 miles	159	0	676	Charleston, WV
250-500 miles	499	4	2,030	Provo-Orem, UT

Notes: Descriptive statistics taken from USGS database.

App. Table 2b: Descriptive statistics on coal deposits, present

	Mean	Minimum	Maximum	Maximum City
Total coal deposit counts				
Total 0-100 miles	100	0	1,492	Huntington-Ashland, WV-KY-OH
Total 0-250 miles	623	0	3,374	Parkersburg-Marietta, WV-OH
Total 0-500 miles	2,077	0	4,994	Louisville, KY-IN
100-250 miles	522	0	3,030	Dayton-Springfield, OH
250-500 miles	1,454	0	4,584	Terre Haute, IN

Notes: Descriptive statistics taken from USGS database.

App. Table 3a: Sample decomposition of IV estimations for average establishment size with 500-mile band

	Total employment growth	Mining, construction & mfg. sectors	Trade, finance, & services sectors	Level of industry agglomeration			High agglom. trade, finance, & services
				Low	Medium	High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Measuring entrepreneurship through average establishment size							
Log average establishment size in city at start of period	-0.878 (0.251)	-0.481 (0.287)	-0.696 (0.271)	-0.502 (0.251)	-0.593 (0.239)	-1.778 (0.346)	-0.942 (0.330)
First stage statistics	Partial R squared 0.193; F statistic 0.000; Maximum 2SLS relative bias <10%						
Over-identification test p-value	0.357	0.656	0.272	0.207	0.816	0.707	0.181
Exogeneity test p-value	0.281	0.748	0.186	0.507	0.231	0.001	0.068
B. Panel A with warm/cold city interaction							
Log average establishment size in city at start of period	-0.899 (0.218)	-0.506 (0.297)	-0.731 (0.235)	-0.533 (0.296)	-0.613 (0.263)	-1.791 (0.354)	-0.995 (0.281)
x indicator variable for warm city (January temperature >34)	0.062 (0.055)	0.074 (0.090)	0.105 (0.066)	0.093 (0.077)	0.059 (0.071)	0.038 (0.135)	0.158 (0.133)
Linear combination for warm cities	-0.836 (0.230)	-0.432 (0.284)	-0.626 (0.250)	-0.440 (0.309)	-0.554 (0.254)	-1.753 (0.342)	-0.837 (0.300)
First stage statistics	Shea partial R squared 0.197 and 0.070; Maximum 2SLS relative bias <20%						
Over-identification test p-value	0.264	0.609	0.362	0.244	0.974	0.437	0.297
Exogeneity test p-value	0.581	0.197	0.421	0.789	0.567	0.006	0.084

Notes: See Table 7. Outcome variables are log employment growth in city as indicated by column headers. The instruments are the log count of mines within 500 miles, 1900; a (0,1) indicator variable for coal or iron ore being the top mineral product in the state, 1928; and the count of iron ore mines within 100 miles of the city, 1900.

App. Table 3b: Sample decomposition of IV estimations for start-up employment share with 500-mile band

	Total employment growth	Mining, construction & mfg. sectors	Trade, finance, & services sectors	Level of industry agglomeration			High agglom. trade, finance, & services
				Low	Medium	High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Measuring entrepreneurship through start-up employment share							
Log start-up share of employment in city at start of period	0.352 (0.119)	0.186 (0.113)	0.260 (0.120)	0.179 (0.121)	0.236 (0.099)	0.761 (0.157)	0.344 (0.160)
First stage statistics	Partial R squared 0.157; F statistic 0.000; Maximum 2SLS relative bias <15%						
Over-identification test p-value	0.083	0.475	0.045	0.070	0.475	0.290	0.036
Exogeneity test p-value	0.024	0.315	0.078	0.285	0.243	0.000	0.082
B. Panel A with warm/cold city interaction							
Log start-up share of employment in city at start of period	0.454 (0.148)	0.267 (0.156)	0.378 (0.128)	0.277 (0.146)	0.315 (0.136)	0.903 (0.179)	0.516 (0.179)
x indicator variable for warm city (January temperature >34)	-0.108 (0.062)	-0.085 (0.070)	-0.125 (0.047)	-0.103 (0.046)	-0.083 (0.071)	-0.150 (0.107)	-0.181 (0.093)
Linear combination for warm cities	0.346 (0.125)	0.182 (0.123)	0.253 (0.112)	0.174 (0.132)	0.232 (0.110)	0.753 (0.149)	0.335 (0.170)
First stage statistics	Shea partial R squared 0.125 and 0.083; Maximum 2SLS relative bias <20%						
Over-identification test p-value	0.288	0.658	0.371	0.255	0.962	0.651	0.310
Exogeneity test p-value	0.015	0.424	0.037	0.208	0.321	0.000	0.017

Notes: See App. Table 3a.

App. Table 4a: Sample decomposition of IV estimations for average establishment size with 250-mile band

	Total employment growth	Mining, construction & mfg. sectors	Trade, finance, & services sectors	Level of industry agglomeration			High agglom. trade, finance, & services
				Low	Medium	High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Measuring entrepreneurship through average establishment size							
Log average establishment size in city at start of period	-0.608 (0.232)	-0.174 (0.314)	-0.432 (0.247)	-0.324 (0.255)	-0.445 (0.270)	-1.161 (0.386)	-0.316 (0.345)
First stage statistics	Partial R squared 0.175; F statistic 0.000; Maximum 2SLS relative bias <15%						
Over-identification test p-value	0.624	0.259	0.293	0.348	0.701	0.044	0.213
Exogeneity test p-value	0.642	0.397	0.874	0.734	0.600	0.302	0.584
B. Panel A with warm/cold city interaction							
Log average establishment size in city at start of period	-0.624 (0.235)	-0.192 (0.346)	-0.458 (0.231)	-0.347 (0.245)	-0.459 (0.282)	-1.170 (0.468)	-0.355 (0.368)
x indicator variable for warm city (January temperature >34)	0.066 (0.055)	0.075 (0.111)	0.108 (0.063)	0.096 (0.062)	0.060 (0.073)	0.040 (0.152)	0.165 (0.137)
Linear combination for warm cities	-0.557 (0.248)	-0.117 (0.329)	-0.350 (0.243)	-0.251 (0.242)	-0.399 (0.273)	-1.130 (0.469)	-0.189 (0.399)
First stage statistics	Shea partial R squared 0.180 and 0.070; Maximum 2SLS relative bias <20%						
Over-identification test p-value	0.956	0.140	0.821	0.703	0.589	0.013	0.497
Exogeneity test p-value	0.760	0.581	0.739	0.739	0.934	0.631	0.321

Notes: See Table 7. Outcome variables are log employment growth in city as indicated by column headers. The instruments are the log count of mines within 250 miles, 1900; a (0,1) indicator variable for coal or iron ore being the top mineral product in the state, 1928; and the count of iron ore mines within 100 miles of the city, 1900.

App. Table 4b: Sample decomposition of IV estimations for start-up employment share with 250-mile band

	Total employment growth	Mining, construction & mfg. sectors	Trade, finance, & services sectors	Level of industry agglomeration			High agglom. trade, finance, & services
				Low	Medium	High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Measuring entrepreneurship through start-up employment share							
Log start-up share of employment in city at start of period	0.245 (0.124)	0.071 (0.143)	0.151 (0.129)	0.101 (0.119)	0.182 (0.136)	0.549 (0.194)	0.085 (0.191)
First stage statistics	Partial R squared 0.129; F statistic 0.000; Maximum 2SLS relative bias <15%						
Over-identification test p-value	0.238	0.254	0.126	0.183	0.508	0.074	0.155
Exogeneity test p-value	0.376	0.989	0.681	0.921	0.557	0.031	0.706
B. Panel A with warm/cold city interaction							
Log start-up share of employment in city at start of period	0.342 (0.132)	0.142 (0.195)	0.266 (0.112)	0.198 (0.132)	0.262 (0.169)	0.672 (0.227)	0.234 (0.208)
x indicator variable for warm city (January temperature >34)	-0.092 (0.052)	-0.068 (0.075)	-0.109 (0.044)	-0.092 (0.043)	-0.076 (0.081)	-0.117 (0.107)	-0.142 (0.096)
Linear combination for warm cities	0.251 (0.116)	0.074 (0.156)	0.157 (0.101)	0.106 (0.112)	0.186 (0.143)	0.554 (0.190)	0.093 (0.216)
First stage statistics	Shea partial R squared 0.102 and 0.082; Maximum 2SLS relative bias <20%						
Over-identification test p-value	0.811	0.162	0.989	0.606	0.727	0.071	0.560
Exogeneity test p-value	0.235	0.945	0.362	0.609	0.629	0.043	0.228

Notes: See App. Table 4a.

App. Table 5a: Extended controls in IV estimations for average establishment size

	Base employment growth estimation	Including log projected employment growth from 1982 to 2002 due to industry distribution	Including log projected establishment growth from 1982 to 2002 due to industry distribution	Including fixed effects for regions x projected employment growth above average	Including fixed effects for regions x projected establishments growth above average	Including log manufacturing employment growth from 1963 to 1981	Including log manufacturing establishment growth from 1963 to 1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. OLS relationship using average establishment size							
Log average establishment size in city at start of period	-0.693 (0.082)	-0.642 (0.109)	-0.690 (0.106)	-0.681 (0.104)	-0.715 (0.090)	-0.686 (0.097)	-0.685 (0.108)
B. IV relationship with 500-mile band							
Log average establishment size in city at start of period	-0.878 (0.251)	-0.698 (0.234)	-0.936 (0.229)	-0.882 (0.230)	-0.873 (0.249)	-0.855 (0.270)	-0.883 (0.254)
First stage partial R squared	0.193	0.165	0.135	0.194	0.186	0.182	0.156
F test statistic in first stage	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum 2SLS relative bias	<10%	<15%	<15%	<15%	<15%	<15%	<15%
Over-identification test p-value	0.357	0.218	0.490	0.389	0.293	0.365	0.357
Exogeneity test p-value	0.281	0.771	0.258	0.325	0.433	0.879	0.324
C. IV relationship with 250-mile band							
Log average establishment size in city at start of period	-0.608 (0.232)	-0.399 (0.254)	-0.641 (0.287)	-0.561 (0.218)	-0.550 (0.213)	-0.581 (0.246)	-0.587 (0.246)
First stage partial R squared	0.175	0.148	0.117	0.180	0.183	0.165	0.149
F test statistic in first stage	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum 2SLS relative bias	<15%	<15%	<20%	<15%	<15%	<15%	<15%
Over-identification test p-value	0.624	0.382	0.432	0.582	0.587	0.578	0.543
Exogeneity test p-value	0.642	0.229	0.867	0.385	0.304	0.579	0.637

Notes: See Tables 2 and 7. Outcome variables are log employment growth in city. The instruments are the log count of mines within 250 or 500 miles, 1900, as indicated; a (0,1) indicator variable for coal or iron ore being the top mineral product in the state, 1928; and the count of iron ore mines within 100 miles of the city, 1900.

App. Table 5b: Extended controls in IV estimations for start-up employment share

	Base employment growth estimation	Including log projected employment growth from 1982 to 2002 due to industry distribution	Including log projected establishment growth from 1982 to 2002 due to industry distribution	Including fixed effects for regions x projected employment growth above average	Including fixed effects for regions x projected establishments growth above average	Including log manufacturing employment growth from 1963 to 1981	Including log manufacturing establishment growth from 1963 to 1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. OLS relationship using start-up employment share							
Log start-up share of employment in city at start of period	0.161 (0.056)	0.134 (0.042)	0.154 (0.056)	0.141 (0.037)	0.168 (0.033)	0.156 (0.052)	0.151 (0.046)
B. IV relationship with 500-mile band							
Log start-up share of employment in city at start of period	0.352 (0.119)	0.266 (0.124)	0.346 (0.132)	0.346 (0.114)	0.341 (0.126)	0.348 (0.142)	0.357 (0.147)
First stage partial R squared	0.157	0.125	0.112	0.158	0.149	0.139	0.116
F test statistic in first stage	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum 2SLS relative bias	<15%	<15%	<20%	<15%	<15%	<15%	<20%
Over-identification test p-value	0.083	0.060	0.092	0.066	0.058	0.082	0.081
Exogeneity test p-value	0.024	0.175	0.059	0.034	0.049	0.037	0.048
C. IV relationship with 250-mile band							
Log start-up share of employment in city at start of period	0.245 (0.124)	0.135 (0.126)	0.229 (0.142)	0.216 (0.107)	0.212 (0.106)	0.235 (0.121)	0.237 (0.133)
First stage partial R squared	0.129	0.100	0.095	0.148	0.147	0.116	0.100
F test statistic in first stage	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum 2SLS relative bias	<15%	<20%	<20%	<15%	<15%	<20%	<20%
Over-identification test p-value	0.083	0.195	0.187	0.263	0.273	0.233	0.222
Exogeneity test p-value	0.376	0.991	0.460	0.582	0.611	0.440	0.448

Notes: See App. Table 5a.

App. Table 6: Quantile instrumental variable regressions

	Mean	Quantiles of the conditional growth distribution					p-value for constant effects
		0.15	0.25	0.5	0.75	0.85	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Measuring entrepreneurship through average establishment size -- quantile regression							
Log average establishment size in city at start of period	-0.693 (0.087)	-0.701 (0.152)	-0.729 (0.142)	-0.820 (0.099)	-0.819 (0.108)	-0.633 (0.116)	0.725
B. Panel A with IVQR using 500-mile band instruments							
Log average establishment size in city at start of period	-0.878 (0.269)	-0.769 (0.449)	-1.095 (0.487)	-1.114 (0.490)	-1.118 (0.383)	-1.121 (0.476)	0.866
C. Panel A with IVQR using 250-mile band instruments							
Log average establishment size in city at start of period	-0.608 (0.232)	-1.074 (0.486)	-0.802 (0.471)	-0.983 (0.559)	-0.648 (0.592)	-0.658 (0.541)	0.715
D. Measuring entrepreneurship through start-up employment share -- quantile regression							
Log start-up share of employment in city at start of period	0.161 (0.040)	0.119 (0.055)	0.158 (0.057)	0.195 (0.079)	0.247 (0.085)	0.240 (0.080)	0.185
E. Panel D with IVQR using 500-mile band instruments							
Log start-up share of employment in city at start of period	0.352 (0.119)	0.360 (0.217)	0.436 (0.198)	0.362 (0.232)	0.375 (0.192)	0.311 (0.212)	0.846
F. Panel D with IVQR using 250-mile band instruments							
Log start-up share of employment in city at start of period	0.245 (0.099)	0.271 (0.237)	0.430 (0.214)	0.338 (0.241)	0.287 (0.296)	0.183 (0.294)	0.963

Notes: See Table 7. Outcome variable is the conditional log employment growth for cities at various quantiles. Growth rates are estimated conditional on the Initial employment controls, Census division fixed effects, and City growth covariates. The instruments are the log count of mines within 250 or 500 miles, 1900; a (0,1) indicator variable for coal or iron ore being the top mineral product in the state, 1928; and the count of iron ore mines within 100 miles of the city, 1900. Instruments are indicated by panel titles. Regressions are unweighted, report bootstrapped standard errors, and have 291 observations. Column 7 reports the p-value from a Wald test where the null hypothesis is that the effects at the 15th and 85th quantiles are equal.