Labor Market Density and Increasing Returns to Scale: How Strong is the Evidence?

Yu-chin Chen (University of Washington)
Noah Weisberger (Goldman, Sachs & Co.)
Edwin Wong (University of Washington)

February 2011

Abstract. Models of economic geography posit that the density of economic activity has two effects that oppose each other in equilibrium: decreasing returns to productive activities due to congestion effects and increasing returns that result from information spillovers and local demand externalities. In an influential paper, Ciccone and Hall (1996) looked at the effect of county level labor market concentration on per-worker Gross State Product in a cross section of US States, and observed that on net, the increasing returns/agglomeration effect dominates. We extend their analysis and re-examine the relationship between density and productivity across industries and over both states and time. Through careful identification of the source and nature of productivity shocks, we show that the evidence for agglomeration effects is indeed quite robust, even within industries, providing evidence for the presence of Marshallian externalities. As for the balance of agglomeration and congestion effects found in previous literature, what we call "net increasing returns to scale", the evidence is much weaker.

J.E.L. Codes: R12, O4
Key words: geographic density; labor productivity; increasing returns to scale

Acknowledgements. We thank Christopher L. Foote for his guidance and provision of data. We are also grateful to Gordon Hanson for sharing his county-reclassification codes and John C. Williams, Yannis M. Ioannides and the two anonymous referees for their helpful comments on earlier drafts of this paper. Feedback from participants at the Harvard University macroeconomics workshop are also acknowledged. Bisundev Mahato and Tripti Thapa provided excellent research assistance. All errors, however, are our own.

"We should be justifiably skeptical of any test of agglomeration that did not control for geographic concentration due to the supply of exogenous site-specific resources."

Ellison and Glaeser (Journal of Political Economy, 1997)

1 Introduction

Theoretical models of economic geography posit that the geographic concentration of economic activity has two effects that oppose each other in equilibrium: decreasing returns to productive activities due to congestion effects and increasing returns that result from information spillovers and local demand externalities.[1] This paper explores how spatial agglomeration of production at the local level influences aggregate labor productivity of the region. The heterogeneity across US states, in terms of labor productivity and the density of economic activity, allows us to test for the presence of increasing returns to scale that may exists across US states and over time. Using the density of economic activity as our measure of geographic concentration, we specifically examine the existence and strength of spatial agglomeration effects in production both at the aggregate and industry level. A comprehensive examination by industry using density, as opposed to size to measure geographic concentration, to our knowledge has not been fully addressed.

The concept of external economies, first illustrated by Alfred Marshall (1920), has occupied a central role in urban economics, as it is closely allied with the observed spatial concentration of production. Fujita, Krugman, and Venables (1999) provide a comprehensive dynamic general equilibrium framework that addresses the channels and forces of agglomeration, pricing and factor market dynamics, and the stability of equilibria.

The vast majority of the evidence in literature supports the hypothesis of agglomeration

[1] Krugman (1991) presents the canonical theory of economic geography. For a comprehensive discussion, see Fujita, Krugman and Venables (1999).

economies.[2] Despite the evidence, empirical identification and quantification of the relevant equilibrium forces are difficult tasks. As such, we follow many previous studies that look at density as the key determinant of productivity by relying on partial equilibrium analysis.[3]

In an influential paper, Ciccone and Hall (1996) model the externality from local production as arising from the density, rather than the size of economic activity, and examine the effect of county level labor market concentration on Gross State Product per worker in a cross section of US States.[4] They estimate, via non-linear instrumental variables, the returns to scale of these local economies as the net effect of (neoclassical) congestion effects on the one hand, and agglomeration effects on the other. Instrumental variables were used to address possible reverse causality in the relationship between labor productivity and spatial density and included historical data for size and density, as well as characteristics that affected population in the past. They find evidence of net increasing returns to scale due to spatial agglomeration. According to Ciccone and Hall, the doubling of average state density increases average state labor productivity by about six percent.[5]

Our work builds on the original work of Ciccone and Hall and explores the robustness of their findings. More specifically, we use variation across both states and time to better

[2] In a survey of empirical studies, Rosenthal and Strange (2004) find that most studies find that doubling city size is associated in a 3-8% increase in labor productivity.

[3] Under a partial equilibrium framework, we allow for the possibility that net local economies of scale can exist in equilibrium. The focus of our analysis is on estimating the scale of local economies, while putting aside the question of local factor prices. Clearly, in order for net local economies of scale to exist in equilibrium, there must be other factors, such as higher prices of housing or other local amenities, that drive a wedge between the production and consumption wage, hence eliminating the incentive for workers and firms to relocate.

[4] Ciccone and Hall (1996) use county level employment relative to physical space (acres) to capture the degree of agglomeration within a state. They argue that density, rather than size, is a more accurate measure of the extent of agglomeration and spatial externalities. They also choose to look at density relative to physical space because it is less subject to the (perhaps) arbitrary classification of "city boundaries." Other studies on the size of industries, such as Henderson (1986), do not focus specifically on the localization of the external economies.

[5] Later studies have confirmed the existence of agglomeration effects. For example, Ciccone (2002) find agglomeration effects in Europe that are slightly smaller than analogous effects in the United States. Combes et al. (2010) also added geological variables as instruments, and confirm the existence of agglomeration economies in French data. Ke (2009) use a simultaneous equation model and find that the spatial concentration of industrial production is an important cause of higher productivity across Chinese cities.

control for state fixed effects and state-level business cycle fluctuations in attempt to isolate exogenous movements in density.[6] In addition, we explore the extent to which differences in industrial composition across states may play in explaining the observed correlation between production density and labor productivity.

The thought experiment that implicitly must stand behind any econometric identification of the effects of labor density on productivity is that density is assigned exogenously to US counties. Clearly the identification issue looms large. Is New York City productive because it is densely populated or is it densely populated because economic activities are especially productive there? Our test of the robustness of the Ciccone and Hall finding seeks to address a number of identification issues that are of potential concern.

Firstly, we estimate the density-productivity relation at the industry level. To our knowledge, a detailed comparison of agglomeration effects across industries has not been address in the current literature. This analysis is motivated in part by Costello (1993), who highlights the role industry specific effects play in explaining aggregate productivity fluctuations. We address the possibility that the positive net increasing returns to scale found by Ciccone and Hall may be driven by particular industries within a state. Suppose high productivity industries locate in dense states and low productivity industries locate in low density states, then a possible interpretation for the finding of increasing returns to density based on cross state observations is that it merely reflects the particular industry mix within a state. In other words, states specializing in highly productive industries may, by necessity, be dense, whereas low productivity states may be less spatially dense, and cross

[6] The finding of increasing returns in local economies, where agglomeration effects form the channel for the externality, has been confirmed by Harris and Ioannides (2000). They look at metropolitan statistical area real wages and population density, as opposed to state level productivity and the density of employment. Their data comes from the decennial census, which covers every 10th year from 1950 to 1990. Harris and Ioannides do not attempt to control for secular trends over the 50 years of data that they use. And their estimations are done via ordinary least squares, without controlling for the potential endogeneity of density with respect to productivity/wages. Because our interest is to provide a more careful and detailed examination of Ciccone and Hall (1996), we followed their state-level approach.

state comparisons do not capture measures of local externalities.

Additionally, industry-level analysis may help distinguish between two potential sources of economies of scale, namely the so-called Marshallian externalities, which arise from firms in the same industry locating near one another, and urbanization externalities, which are the positive effects that arise from the general diversity and scale of urban areas. Studies which investigate industry specific effects have focused primarily on the manufacturing sector. For example, Wheeler (2006) find that the increasing relationship between workers wages and average plant size. Ellison et al. (2010) find that labor pooling, knowledge spillovers and input-output linkages are all important in explaining conglomeration patterns among manufacturing industries. Glaeser and Gottlieb (2009) also conclude that input-output linkages are important for manufacturing firms, but is more important for service firms. Nevertheless, they note that agglomeration operating though knowledge spillovers and the transfer of information are more important. Kolko (2000) argues that advances of information technology have diminished the importance of transaction costs in the location of firms in the service industry. Instead firms tend to locate near workers that the industry demands.

Secondly, the results of Ciccone and Hall and our thought experiment above suggest that were we to uproot a worker from the middle of Idaho to Manhattan, the worker's productivity would rise purely because New York city is denser. This story assumes that there are no New York specific effects that cause economic activity to be highly productive, which also has encouraged people to locate there. As pointed out by Hanson (2000), Ellison and Glaeser (1997), among others, regional amenities are likely to simultaneously affect labor productivity (or wages) and the location decision of workers, i.e. density. To the extent that geographic amenities are fixed in time and place, say for example New York Harbor, we employ panel estimation with state fixed effects to control for this potential source of bias. Henderson (2003) employ a similar approach by including city and time fixed effects to identify scale externalities for the machinery and high tech industries. Although Ciccone and Hall attempt to control for the endogeneity of density by using a variety of instruments (all of which can be loosely thought of as lagged density), if past density and current productivity are both impacted by the same state fixed effect, then cross sectional estimates would still be biased. In this case, panel estimation would be necessary and may in fact, be all that is needed.[7]

Thirdly, cyclical fluctuations in technology may simultaneously affect both labor productivity and employment density. As discussed in Blanchard and Katz (1992) and Foote (2000), geographic labor flows are quite sensitive to business cycle conditions. If people move to high productivity places in response to good technology outcomes and low unemployment rates, we would be identifying the effect of density on productivity off of endogenous cyclical variation. In the panel estimation discussed above, such state cycles hamper identification.[8] To address this, we use measures of local aggregate demand to control for state business cycle effects.

Even after controlling for cyclical effects as well as state and time effects, there may still be an endogeneity problem if workers relocate in respond to contemporaneous idiosyncratic productivity shocks. We therefore attempt to isolate truly exogenous movements in density using two alternative instruments. The first instrument we use is lagged density. Although the direction of causality between current labor productivity and current employment density may be blurred, we argue that conditional on state, time, and business cycles, lagged employment density is exogenous with respect to current labor productivity.

As a robustness check, we also use state differences in birthrates 20 and 25 years ago to instrument for current density. In accordance with the stylized fact that birthrates tend to be high in poorer regions and relative state income positions are quite persistent,

[7]Ciccone and Hall use four separate instruments for state density: distance to the Eastern seaboard, existence of a railroad in 1860, population in 1850, and population density in 1880.

[8]The business cycle fluctuations that we are addressing here are state-specific. Any aggregate US-wide shock can be controlled for in the panel using a time fixed effect. The most readily available state-level cyclical control is state unemployment rates. But as detailed by Foote (2001), using the unemployment rate in our context would not help us to sort out the causation issue.

we recognize that this instrument may lead to a downward bias in our estimate of the net increasing returns to scale. Although far from perfect, under certain assumptions about workers behavior, these two instruments allow us to treat the resulting movements in density as exogenous with respect to the state and time specific productivity shocks, or at least, provide us with an estimated range of the effect of density on average labor productivity.

Previewing our results below, through careful identification of the source and nature of productivity shocks, we show that the evidence for agglomeration effects is quite robust, even within industries. This suggests the presence of Marshallian externalities within many industries. Moreover, industry specific estimates point to the presence of spillover effects across industries. As for the balance of agglomeration and congestion effects, what we call "net increasing returns to scale," the evidence is mixed.

The rest of the paper proceeds as follows. Section 2 presents the theoretical framework underlying our analysis. Section 3 describes the data and illustrates some of the basic empirical facts about state and county level economic activity. Section 4 presents the empirical findings, and in section 5, we discuss our findings and conclude.

2 Theoretical Framework

Increasing returns from the density of production may arise from many sources, either external or internal to the firm. Marshall (1920) and Henderson (1974, 1988), for example, focus on industry-specific technology and knowledge spillovers between nearby firms that induce firms to agglomerate. Knowledge spillovers have been verified empirically by Henderson (2003), Rosenthal and Strange (2001) and Jaffe et al (1993) among others. Fujita (1988, 1989) and Krugman (1991) utilize the Spence-Dixit-Stiglitz monopolistic competition market structure to demonstrate demand linkages as the source of increasing returns. When trade is costly (for example iceberg type transportation costs), easier access to a greater variety of local intermediate inputs and services provide incentives for firms to locate near larger markets. Duranton and Puga (1999) and Lucas (1988) offer examples of how localized human-capital externalities may serve as forces of agglomeration. Acting against these local externalities are congestion costs associated with limited local supplies of non-traded goods and other negative amenities associated with spatial crowding. Other sources of increasing returns that are highlighted in literature include input sharing (Burchfield et al., 2006; Holmes, 1999), labor market pooling (Overman and Puga, 2010; Ellison et al., 2010), search and matching process (Gan and Li, 2004; Costa and Kahn, 2000; Helsey and Strange, 1991) and establishment scale (Wheeler, 2006).

Estimation of general equilibrium models of external economies, however, need to confront the possibility of multiple equilibria, as discussed in Hanson (2000). Instead, we follow the partial equilibrium approach formulated by Ciccone and Hall, where geographically localized externalities in production at the county level are the source of aggregate increasing returns. And congestion effects take the form of neoclassical decreasing returns to inputs.

The main difference between our model and the model of Ciccone and Hall is that we specifically incorporate time into the model and include a more detailed treatment of the technology shocks across time and geography. The way in which we specify the technology shock will provide the identification assumptions in our estimations, as will be seen below in Section 4.

In our model, production is assumed to take place at the county level on each acre of land. Let y_t denote output on an acre of land. The inputs to production are labor per acre, denoted n_t, which is augmented by h_t, the efficiency of labor, and capital per acre, which is denoted by k_t. Let Λ_{st} denote a Hicks neutral production technology that is the same everywhere within a state s at time t. The agglomeration externality is modeled by linking production in the county as a whole to per acre production for all acres in the county. Specifically, output per acre is influenced by production per acre in the county as a whole, represented by $\frac{y_{ct}}{a_c}$, where y_{ct} is county-wide output at time t, and a_c denotes the total acreage in county c. Output per acre is assumed to take the following form:

$$y_t = \Lambda_{st}[(h_t n_t)^{\beta} k_t^{1-\beta}]^{\alpha} \left(\frac{y_{ct}}{a_c}\right)^{\frac{\lambda-1}{\lambda}}. \tag{1}$$

The labor and capital inputs are governed by the neoclassical decreasing returns to scale parameters $\alpha\beta$ and $\alpha(1-\beta)$, where α is smaller than one by the amount the share of land occupies in total factor payments (typically about 5 to 6 percent). The output externality from economic activity is a function of the density of economic activity at the overall county level, with constant elasticity $(\lambda-1)/\lambda$. The existence of localized production externalities would imply $\lambda > 1$.

Assuming inputs are evenly distributed across all acres within a county, per acre variables can now be substituted with county level variables (i.e. $y_t = \left(\frac{y_{ct}}{a_c}\right)$, $n_t = \left(\frac{n_{ct}}{a_c}\right)$,and $k_t = (\frac{k_{ct}}{a_c}))$, and we can solve for output per acre:

$$\frac{y_{ct}}{a_c} = \Lambda_{st}^{\lambda}[(\frac{h_{ct}n_{ct}}{a_c})^{\beta}(\frac{k_{ct}}{a_c})^{1-\beta}]^{\gamma} \tag{2}$$

where

$$\gamma = \alpha\lambda \tag{3}$$

γ captures the competing effects of the decreasing marginal product of factors in α, which can be thought of as a congestion effect, and the agglomeration effect of λ. Values of $\gamma \geq 1$ imply that the effects of agglomeration outweigh the effects of congestion, and on net, higher density of economic activity leads to an increase in total factor productivity. As a benchmark, we treat α to be around 0.95, with land occupying about 5% of total factor payments.[9] Any estimate of γ greater than 0.95 would then imply some degree of

[9] Ho, Jorgenson and Stiroh (2002) show that land's share of factor payment in the US economy is between

agglomeration effects, i.e. $\lambda > 1$, and that the agglomeration effects outweigh the congestion effects on net.

Assuming perfect capital mobility, the rental price of capital r_t will be the same everywhere at each time t. Using the factor demand equation, we can substitute $k_{ct} = \alpha(1-\beta)y_{ct}/r_t$, and arrive at the following county level output equation:

$$y_{ct} = c_t \Lambda_{st}^{\omega} [a_c (\frac{h_{ct} n_{ct}}{a_c})^{\theta}] \tag{4}$$

where c_t is a function of the interest rate at time t and invariant across state. The elasticity for the state technology multiplier ω is

$$\omega = \frac{\theta}{\alpha\beta} \tag{5}$$

and the elasticity of county employment density θ is

$$\theta = \frac{\gamma\beta}{1-\gamma(1-\beta)}. \tag{6}$$

Because reliable measures of output are available only at the state level, we sum up county output across all counties in a state (we use capital letters denote state level variables). We further assume that labor quality h_{ct} is the same within all counties of a state and depends log-linearly on worker's average years of education, ed_{st}. State level output can be expressed as

$$Y_{st} = c_t \Lambda_{st}^{\omega} ed_{st}^{\eta} \sum_{c \in S} a_c \left(\frac{n_{ct}}{a_c}\right)^{\theta}. \tag{7}$$

Dividing through by total state employment and taking logs on both sides, state level output

5 and 6%.

per worker can be expressed as:

$$\ln\left(\frac{Y_{st}}{N_{st}}\right) = \ln(c_t) + \omega\ln(\Lambda_{st}) + \eta\ln ed_{st} + \ln\left(\frac{\sum_{c\in S} a_c^{-(\theta-1)}(n_{ct})^{\theta}}{N_{st}}\right). \tag{8}$$

In other words, state level labor productivity is a function of employment density at the county level.

The two key parameters are θ, the returns to employment density, and $\ln(\Lambda_{st})$, state level technology. Similar to our discussion above, estimates of θ greater than 1 imply that there are net increasing returns to scale as agglomeration effects outweigh congestion effects. And as a benchmark, estimates of θ greater than about 0.92 imply that, even if on net neoclassical congestion effects dominate, agglomeration effects are still important ($\lambda > 1$).[10]

The way in which we specify $\ln(\Lambda_{st})$ will help clarify our identification assumptions in the empirical work below. For example, Ciccone and Hall assume that state technology is log-normally distributed around a country-wide time invariant mean. So in their cross section estimation, $\ln(\Lambda_{st})$ reduces to $\psi + \varepsilon_s$. In our most general specification, we assume that shocks to $\ln(\Lambda_{st})$ to have a state specific term, a time specific term, a cyclical term, and a term that is iid and can be thought of as a purely random, one-off innovation to technology that is uncorrelated across time or states.

3 The Data and Some Facts

3.1 The Data

The data used in this paper are annual from 1982 to 1999 for the District of Columbia and all states except Alaska.[11] Our measure of output at the state level is real Gross State

[10] Section 4.1 below explains how the 0.92 benchmark is derived.
[11] We maintain the exclusion of Alaska that was employed in the original Ciccone and Hall paper.

Product (GSP) and comes from the Bureau of Economic Analysis. We use both total real GSP (1996 dollars) by states as well as real GSP by state by one-digit Standard Industrial Classification (SIC) industry. The chained price that is used to deflate GSP data is based on the national price of goods produced in each industry.[12] The aggregate state real GSP and state real GSP by one-digit SIC industry are chain-weighted sums of the disaggregated pieces. Real state GSP represents the composition of state production valued at common national prices. Differences in real GSP across states reflects differences in the production bundle of the states evaluated at a common price. Cross-state differences in input and/or output prices are not reflected in these data.

Employment by one-digit SIC industry by county is available from the Bureau of Economic Analysis Regional Economic Information System CD. We aggregate up the county level employment to get state employment both in total and by industry. And we define labor productivity as real state output divided by state employment.

Lagged employment density, which is used as a state level instrument in the estimation of (8) is constructed as an index measure of the average county level employment density within a state:

$$density_{st} = DI_{st} = \sum_{c \in s} \left(\frac{n_{ct}}{N_{st}} \right) \left(\frac{n_{ct}}{a_c} \right) \tag{9}$$

where s indexes US states, c indexes the counties of state s, N_{st} is employment in state s, n_{ct} is employment in county c at time t, and a_c is the number of acres in county c. This index is a linearized version of the measure of local employment density in equation (8).[13]

[12] The BEA has price data for 63 industries. Real state GSP is generated by summing up (using chain-weights) all 63 components of state output. For a more detailed discussion see the August 2001 *Survey of Current Business.*

[13] In this paper all estimation is done via non-linear least squares as dictated by the functional form of equation (8). In unreported results, we estimated alternative specifications with ordinary least squares using this linearized index in place of the last term in (8). The results are quite similar to those reported in this paper. We also used the linear specification to check the explanatory power of all instrumental variables. Instrument relevance was confirmed in all specifications for lagged density, and the majority of specifications for birthrates.

Human capital is measured as the hours-weighted average years of schooling for workers in each state by industry and is based on data from the monthly Current Population Survey. In 1992, the CPS switched from asking respondents about years of schooling to asking them about educational attainment. Following Jaeger (1997), we map educational attainment data from post-1992 CPS data to years of schooling by using an imputed highest grade completed, which attempts to make the two parts of the series consistent.

3.2 Some Facts

Labor productivity is extremely heterogenous across states and across industries. In 1982, real average productivity (output per worker in 1996 dollars) was about 41,000 dollars for the US as a whole (see Table 1). By 1999, this figured had grown about 25 percent to just over 51,000 dollars. However, this growth conceals a good deal of the underlying heterogeneity. The standard deviation of state-level productivity was about 5,000 dollars in 1982 and had grown to over 8,000 dollars in 1999. To better illustrate the cross state differences in productivity, take for example New York and Montana. New York led the nation in labor productivity in 1999, at over 70,000 dollars per worker. In contrast, an average Montana worker produced only 36,000 dollars of output, nearly half the level of a New York worker, and had the lowest productivity in the country. In 1982, New York ranked second in the nation in labor productivity at just over 51,000 dollars per worker. Montana was ranked in the middle of the country in 1982 with productivity of nearly 39,000 dollars per worker. In other words, from 1982 to 1999, average (across states) productivity grew about 25 percent, New York's productivity grew over 36 percent, and productivity in Montana *fell* by about seven percent.

Turning to density, the least dense state in 1999 was Wyoming with less than one worker per hundred acres. Not surprisingly, the most dense state is New York with over 41 workers per acre. Density by industry varies quite a bit as well. Columns 4 and 7 of Table 1 present summary statistics for the density index constructed using expression (9) by industry. Because the index weighs a county's employment density by the population density, counties with a sparse population contribute less to the state level index than a densely populated county would. For example, the mean employment density index of 37.27 for the finance, insurance and real estate sector in 1982 reflect the fact that the high employment density in New York City is weighted more heavily than the low employment density in rural Wyoming. Total density has grown over time from 192.39 workers per acre in 1982 to 207.34 workers per acre in 1999. Average density is greatest in the services sector with 62.48 workers per acre in 1982, followed by government with 42.16 workers per acre. In contrast, agriculture has, on average, 0.75 workers per acre.

With regard to the composition of state output in Table 2, we focus on manufacturing and services. Clearly, these two industries have experienced vastly different fortunes. Manufacturing has declined in importance, with its share of employment and output falling over time. However, the labor decline has outpaced the decline in its share of output, so manufacturing's productivity has increased by over 66 percent. Not only has the average level of productivity increased, the standard deviation has more than doubled as well. In contrast to manufacturing, services has seen modest gains in employment share, output share, and labor productivity.

4 Identification, Estimation and Results

Spatial increasing returns are difficult to observe or quantify empirically, providing researchers little choice but to make inference indirectly. As summarized in Hanson (2000), empirical identification of agglomeration economies is a task wrought with difficulties, with three key complications that are particularly difficult to resolve: unobserved regional characteristics, simultaneity in regional data, and multiple sources of externalities. Although we are far from claiming our approach has resolved these difficulties, the estimation procedure we discuss below allows us to more systematically evaluate the relevance of each.

Taking equation (8) as our starting point

$$\ln\left(\frac{Y_{st}}{N_{st}}\right) = \ln(c_t) + \omega\ln(\Lambda_{st}) + \eta\ln ed_{st} + \ln\left(\frac{\sum_{c\in S} a_c^{-(\theta-1)}\left(n_{ct}\right)^{\theta}}{N_{st}}\right),$$

we will make explicit assumptions about $\ln(\Lambda_{st})$, which will give us guidance as how best to estimate θ.

4.1 Controlling for Industry Composition

We first follow the assumptions made by Ciccone and Hall that at each time t, $\ln(\Lambda_{st})$ is log-normally distributed around a nationwide mean. It can then be expressed as $\psi + \varepsilon_{st}$ where ε_{st} is uncorrelated across states. Therefore, θ can be estimated using cross state variation. Dropping the time subscripts, equation (8) simplifies to

$$\ln\left(\frac{Y_s}{N_s}\right) = \psi + \ln\left(\frac{\sum_{c\in S} a_c^{-(\theta-1)}\left(n_c\right)^{\theta}}{N_s}\right) + \eta\ln ed_s + \varepsilon_s. \tag{10}$$

which we estimate using data from 1999 via non-linear least squares.[14]

Again, values of $\widehat{\theta} > 1$ (where the hat denotes the estimate of θ) indicate that on net, the effect of agglomeration dominates the effect of congestion, or that the externalities of production density outweigh the neoclassical decreasing marginal product of factors.

The first panel of Table 3 presents the results both for total state output per worker as well as output per worker by one-digit SIC industries. For total GSP, the estimated θ is 1.08 and is significantly greater than 1. This implies that a doubling of employment

[14] We performed the same estimations for other years as well but do not report them here. The results are qualitatively similar to the ones for 1999.

density increases average state labor productivity by 8 percent. Log education does not enter significantly in this base line specification. Our initial estimates are in line with the results of Ciccone and Hall.

To understand the implications of $\theta = 1.08$, we use the following benchmark factor payment decomposition: 5% land ($\alpha = 0.95$), 60% labor ($\alpha\beta$), and 35% capital ($\alpha(1-\beta)$). From equations (5) and (6) above, $\theta = 1.08$ implies that γ is around 1.05.[15] That is, doubling the production density in a county results in a 5% increase in total factor productivity. The implied agglomeration parameter λ is roughly 1.10, suggesting that the elasticity of output with respect to density is roughly 9%. In fact, in our benchmark economy, any $\widehat{\theta}$ larger than 0.92 would imply some degree of production economies of scale ($\lambda > 1$) even if on net, agglomeration is dominated by congestion.

Motivated by our summary statistics in Tables 1 and 2 which suggest that output per worker looks very different by industry, we estimate θ separately for each industry, again using non-linear least squares. By looking across industries, we can see if industries differ significantly in their respective returns to scale, and whether failure to control for state differences in industry mix may bias our returns to scale estimates. In addition, we can see if the returns to scale at the state level arise from inter-industry externalities (urbanization) or intra-industry externalities (specialization).

Our estimates do not rule out externalities from specialization and suggest proximity to the same industry does provide some net economies of scale. For six of the nine industries, our estimates of θ are significantly greater than 1. In the remaining three industries, namely agriculture, manufacturing, and transportation and public utilities, we cannot reject the null hypothesis that $\theta = 1$. However, the estimates for θ are all above 0.92, suggesting that agglomeration effects still play an important role, even when we look at productivity within

[15] These calculations are done using equation (5) $\gamma = \frac{\theta}{\beta+\theta(1-\beta)}$ and (6) $\lambda = \gamma/\alpha$.

industries.[16]

4.2 Endogeneity

Up to this point, we have assumed that density is exogenous with respect to productivity in the cross section. We now relax this assumption and re-estimate θ using a generalized method of moments instrumental variables approach. Ciccone and Hall use four different instruments for contemporaneous state employment density: the presence of absence of a railroad in 1860, state population in 1850, state population density in 1880, and distance from the eastern sea board. In all cases, they find $\widehat{\theta} > 1$. All of their instruments seem to reflect or are highly correlated with lagged employment density. Because of the panel set-up that we will discuss below, we need to find an instrument that varies over time. Therefore, we constructed an explicit measure of lagged density. Our instrument is the weighted average of county level employment density in a state, using the employment share of each county as weights, as described earlier.[17] This instrument is valid under the assumption that workers relocate in response to fairly contemporaneous shocks in productivity only, so density from half a decade or more ago is not influenced by current labor productivity.

In our cross sectional estimates we use two different instruments. In panel 2 of Table 3, we use employment density in 1977 to instrument for employment density in 1999, and in panel 3, we use employment density in 1994 as an instrument. The two instruments give very similar results and once again confirm that in the cross section, net increasing returns to the density is present. Looking at both the aggregate and the within industry estimates, the pattern is little changed. Across both choices of instrument, θ is estimated to be 1.08 and significantly greater than one for total GSP per worker. To the extent that workers do

[16] We note that the 0.92 cutoff calculated using 5%-60%-35% factor share decomposition may not be appropriate for all industries, as they differ in their production factor intensity. Here we merely use it as a rough benchmark.

[17] The county density measure we use as an instrument is a linearized version of the county density that appears as a regressor in the NLS estimations, as discussed in section 2.

not move in anticipation of potential high productivity in a state 20 years later, it seems our 5 year lagged density may also be a valid instrument as it provides very similar results.

We note that if productivity shocks are extremely persistent (after removing state specific and business cycle related components), both our instruments as well as Ciccone and Hall's instruments would induce upward bias in the estimate of net increasing returns.[18] Therefore, we consider an alternative instrumental variables in our panel setup, as will be discussed later.

Turning to industry level data, we now do find the estimated net returns to scale for agriculture to be significantly below 1. However, this is not surprising as agriculture is quite land intensive. While the estimated θ's (0.91 using density in 1977 and 0.95 using density in 1994) may imply net decreasing returns relative to the 0.92 benchmark set up above (which attributed little weight to land as a factor of production), the degree of net returns to scale is actually quite large relative to a more realistic benchmark for agriculture.[19] This seems to indicate that even in cases where on net the agglomeration effect does not dominate, it still plays a significant role.

With the exception of agriculture, the instrumental variables estimates of θ for the other industries suggest that agglomeration effects dominate within industry as well. As agriculture's share of production is quite small, and the differences in the estimates for the other industries are not big, we conclude that failure to control for industrial composition differences across state may not cause much bias in the estimations using aggregate data.

[18] High persistence in state specific productivity shocks (after removing state specific and business cycle induced components) would suggest that there is a positive correlation between the instruments (lagged density) and current productivity shocks, which is our regression residual. As density is also positively autocorrelated, the estimate θ would thus be upward biased.

[19] For example, if we assume labor, capital, and land to occupy equal factor shares in agricultural production, any $\theta > 0.38$ would imply positive agglomeration externalities.

4.3 State and Time Fixed Effects: Panel Data

To this point, our identifying assumption in the non-linear least squares estimation has been that workers choose which counties to live in randomly with respect to the average productivity of the state. In our instrumental variables estimation, the identifying assumption is that the choice of location made by workers some number of years past is unrelated to productivity today. Even this second identifying assumption seems problematic. Time invariant state fixed effects may drive both the decision to locate in a particular state as well as productivity in that state. To the extent that positive state amenities may increase productivity and attract workers simultaneously, failure to address this state fixed effect would lead to upward bias in the estimate of θ. Returning to our basic equation (8), we now allow the productivity term, $\ln(\Lambda_{st})$, to be log-normally distributed around a state specific component and a time specific component. The error term in our equation can now be expressed as $\varepsilon_{st} = \upsilon_s + \zeta_t + \mu_{st}$ and we obtain the following estimation equation:[20]

$$\ln\left(\frac{Y_{st}}{N_{st}}\right) = \eta \ln ed_{st} + \ln\left(\frac{\sum_{c\in S} a_c^{-(\theta-1)} \left(n_{ct}\right)^{\theta}}{N_{st}}\right) + \upsilon_s + \zeta_t + \mu_{st}. \tag{11}$$

To the extent that density is correlated with the time invariant state specific term, υ_s, the cross-sectional results are biased. If we assume the entire endogeneity problem in the cross section is due to a state fixed effect that causes productivity and density to be correlated, then the panel fixed effect estimation should remove all of the endogeneity, leaving the remaining variation (arguably) exogenous with respect to density.

We estimate the above equation, where both state and time fixed effects are controlled, using panel non-linear least squares and present the results in the first panel of Table 4. Interestingly, we find that for total GSP, the point estimate of θ is less than 1, which stands in contrast to the cross sectional results, even though we cannot reject the possibility that

[20] The $log(c_t)$ term is absorbed into the state specific ζ_t term.

$\theta \geqq 1$. However, the point estimate is large enough so that even if, on net, congestion effects dominate, the agglomeration externality is still present. Looking at the industry specific estimates of θ, the evidence is a bit more mixed. Construction and manufacturing are the only two industries that have estimated θ that are significantly greater than 1. Agriculture has the lowest estimated θ at 0.77, which is much more in line with a neoclassical production function.

While we feel that adding a state-fixed effect goes a long way toward attempting to control for correlation between density and the error term, we also estimate θ via generalized methods of moments by instrumenting for density with 5 year lagged density in the panel setup. Results from this GMM-IV estimation are given in panel 2 of Table 4. The GMM-IV estimates of θ are more in line with the increasing returns to scale story. For total GSP per worker, θ is estimated at 1.05, only slightly lower than the cross-sectional IV estimate where we did not control for state specific effects.

4.4 Agglomeration Effects by Industry

In line with results for total GSP, estimates of θ by industry after correcting for endogeneity are similar to the NLS measures. However, controlling for state and time fixed effects significantly effects the estimates of θ for many industries. Two industries stand out with lower estimates of θ relative to other industries in the sample. Agriculture, with an estimate of θ less than 1 in all specifications, is land intensive and less likely to benefit from spillover effects that other industries in more urban areas would. Similarly, the transportation and public utilities sector, which primarily supply services to other industries are less likely to benefit from knowledge spillovers, input-output externalities and other sources of agglomeration.

In contrast, agglomeration effects are particularly strong in the manufacturing and construction sectors. The large effects for manufacturing are consistent other studies (Ellison et al., 2010; Glaeser and Gottlieb, 2009, Wheeler, 2006). These two sectors are most likely to benefit from both positive Marshallian and urbanization externalities. For example, both sectors benefit from labor market pooling, input cost reductions and knowledge spillovers that have emphasized as important causes of agglomeration in other studies. Moreover, firms in these sectors, particularly manufacturing also benefit from economics of scale. Agglomeration effects are also present in service oriented sectors that include trade, FIRE and government, although effects are much smaller. While all estimates suggest the presence of agglomeration effects given our parameter assumptions, the effect of congestion dominates after adding fixed effects for wholesale trade, services and government.

Using estimates of θ from the GMM estimation of (11) in the third panel of Table 3, we construct the value added weighted average of the industry agglomeration effect estimates for 1999. We obtain a weighted average of 1.05, which is less than the estimate of θ obtained using total GSP. Aside from estimation noise, this difference could reflect the existence of spillover effects that are not captured in the industry specific estimates.

4.5 Regional Business Cycles

Another possible confounding factor that we consider is that people move to high productivity places in response to good technology outcomes or low unemployment rates induced by the state specific business cycles. If this is the case, then we are identifying the effect of density on productivity off of cyclical variation. Even after controlling for state and time fixed effects, there may be an omitted cycle term included in the error that is correlated with density. We need to include some measure of the regional business cycle in the regression equations to control for this.

We choose to control for the regional business cycle with two state-specific aggregate demand measures. The first is real state level defense spending, which reflect Department of Defense expenditures allocated to respective states. This variable represents cyclical measure of government expenditure that is exogenous to idiosyncratic productivity shocks [21]. The second cyclical control is the so-called "Bartik shocks."[22] Bartik shocks are changes in national employment by industry, weighted by state specific industry weights based on 1980 state level industry output shares. They measure the state level impact of changes in the composition of labor demand along a stable labor supply curve.[23]

In terms of our framework above, we now allow our error process to have a cyclical term that varies by state and time. We include four lags each of the Bartik shocks and of defense spending. Results are presented in column 2 of Table 5. (For ease of comparison, Column 1 reports the base-line fixed effects model using NLS.) Lags of both Bartik and real defense spending enter significantly, and the estimate of returns to scale is further lowered to 0.93. It is also significantly less than 1. Compared to the benchmark of $\theta < 0.92$ which implies no agglomeration effect ($\lambda = 1$), this result suggests that not only do net the neoclassical congestion effects seem to dominate, the evidence for the agglomeration effect is also much weaker than our initial estimation. However, in column 4 we instrument for density with lagged density, and our estimate of θ rises back up to 1.07.

It is interesting to note that when we do not use instruments and rely on state and time fixed effects and cyclical controls alone, estimates of θ are either not significantly different from 1 or are actually below 1 (see Table 5 columns 1 and 2). They are in general lower than the θ estimated via generalized method of moments with instrumental variables (see column 3 and 4). Similar to our discussion in section 4.2, if the non-cyclical part of state specific productivity shocks are quite persistent (after removing the state and time specific components), then using lagged density as an instrument would likely bias upward our θ

[21] Barro (1981) find that both permanent and transitory defense purchases have a significant expansionary effect on real GNP.

[22] See "Who Benefits from State and Local Economic Development Policies?" by Timothy J. Bartik (1991)

[23] Other studies employing Bartik shocks include Glaeser et al. (2006) and Saks (2008).

estimates.

Therefore, as a final robustness check, we consider an alternative instrument set for current density: birthrates lagged 20 and 25 years. To the extent that birthrates across states are exogenous to current productivity, high birthrates in a state 20 some years ago should lead to more exogenous labor entry now, hence higher employment density. Using lagged birthrates as instruments, we estimate the base-line panel specification and the panel specification with cyclical controls. Column 5 shows that without cyclical controls, θ is estimated at 0.61. However, once the cyclical controls are included, $\widehat{\theta}$ rises back up to 1.07. While the 0.61 number may suggest net decreasing returns, we note that in reality, lagged birthrates may be negatively correlated with current productivity, as poorer regions tend to have higher birthrates and state income position may be persistent. Therefore, using lagged birthrates as instruments may cause downward bias in our estimates.

5 Discussion and Conclusion

The goal of this paper is to re-examine the results of Ciccone and Hall (1996), which finds evidence of net increasing returns to the density of economic activity by comparing a cross-section of US states. We argue that there are a number of factors that Ciccone and Hall neglected to consider that may be important in estimating the returns to density. Specifically, the empirical challenge we take up here is to isolate, as much as possible, the effect of exogenous density movements on output per worker. Identification is indeed a thorny issue in this context. To the extent that workers move in response to productivity, a state-specific, time-varying instrument that is correlated with density but unrelated to productivity otherwise, is what is required. Given that such instruments are almost impossible to find, we address the endogeneity problem by making specific assumptions about the source and nature of the error term (productivity shock), and try to use the functional form of our estimation equations to correctly estimate the returns to employment density under these assumptions.

A second goal of this paper is to examine differences in returns to scale by industry. Looking at a cross-section of US states in 1999, we find that estimates are consistent with an increasing-returns-to-scale story across most one-digit SIC industries or, at the very least, give evidence in favor of a large role for agglomeration effects. These evidence is weakest for the agriculture and transportation sectors, which are less likely to benefit from Marshallian externalities that other sectors would. We also find evidence that agglomeration effects spillover across industries.

To better control for state specific amenities, we extend the analysis to a panel data framework and control for both state- and time-fixed effects. And finally, we include measures of aggregate demand to control for business cycle effects. In addition to the inclusion of state, time and cyclical controls, we instrument for density with both lagged density and lagged birthrates.

In our panel framework, controlling for time- and state-fixed effects, we estimate θ to be 0.99 for total GSP, and when we include cyclical controls, our estimate falls to 0.93. From these two point estimates, we are tempted to conclude that on net, there are no increasing returns to scale even though agglomeration effects may still play some role. However, when we instrument for density using either lagged density or lagged birthrates, our estimates of θ climb to values significantly greater than 1.

We estimate net returns to scale as the net effect of agglomeration and congestion. Our results suggest that even in a richer empirical framework, there is still strong evidence in favor of the importance of agglomeration effects. The evidence regarding the presence of net increasing returns to density is more mixed, and we conjecture that an alternative modeling framework may be necessary to produce more conclusive results.

Table 1: Cross State Averages by Industry for 1982 and 1999[1]

	1982			1999		
	Productivity	Edu.	Density	Productivity	Edu.	Density
Total	40.87	12.97	192.39	51.08	13.61	207.34
	(5.86)	(0.44)	(623.67)	(8.21)	(.31)	(628)
Agriculture	20.75	11.12	0.75	28.23	12.04	0.46
	(6.53)	(1.39)	(1.82)	(8.30)	(1.37)	(0.55)
Construction	35.37	11.76	4.15	37.47	12.39	4.85
	(8.42)	(0.72)	(7.00)	(7.20)	(0.43)	(6.60)
Manufacturing	42.38	12.19	19.02	78.38	13.01	11.89
	(10.39)	(0.61)	(43.98)	(27.64)	(0.49)	(21.72)
Transportation	65.47	12.52	13.12	91.30	13.27	10.25
	(7.91)	(0.41)	(40.28)	(15.53)	(0.29)	(21.17)
Wholesale Trade	44.92	12.92	14.03	88.89	13.24	9.37
	(4.83)	(0.41)	(53.26)	(13.53)	(0.30)	(26.71)
Retail Trade	20.51	12.34	17.63	29.49	12.83	19.46
	(2.27)	(0.35)	(36.26)	(3.95)	(0.28)	(38.48)
FIRE	109.14	13.60	37.27	115.12	14.05	37.94
	(14.55)	(0.31)	(182.08)	(35.07)	(0.36)	(175.41)
Services	32.53	14.09	62.48	31.22	14.42	88.67
	(4.67)	(0.42)	(213.41)	(6.62)	(0.31)	(277.68)
Government	42.75	13.42	42.16	43.78	14.27	41.54
	(6.50)	(0.37)	(143.06)	(7.22)	(0.35)	(143.46)

[1] Data are from the Bureau of Economic Analysis, the Current Population Survey and authors' calculations. Productivity, measured as GSP per worker, is in 1000s of 1996 dollars. Education is the average years of schooling weighted by hours worked. Density in this table is a linearized version of the density measure used as a regressor and is the county level employment weighted share of workers per acre (see text for further details). Standard deviations are in parenthesis.

Table 2: Cross State Average Output and Employment Shares by Industry for 1982 and 1999[1]

	1982		1999	
	Output Share	Empl. Share	Output Share	Empl. Share
Agriculture	2.39	4.02	1.35	3.18
	(3.43)	(3.25)	(1.36)	(2.04)
Construction	4.06	4.67	4.47	5.65
	(1.12)	(1.27)	(1.18)	(1.08)
Manufacturing	20.22	16.82	16.12	11.76
	(8.73)	(6.63)	(6.66)	(4.36)
Transportation	9.18	4.93	8.38	4.87
	(1.83)	(0.86)	(2.28)	(0.85)
Wholesale Trade	6.96	5.00	6.91	4.56
	(1.55)	(0.96)	(1.52)	(0.85)
Retail Trade	8.95	15.86	9.19	16.43
	(1.44)	(1.59)	(1.52)	(1.77)
FIRE	15.56	7.75	19.25	7.92
	(3.06)	(1.38)	(5.79)	(1.62)
Services	14.75	23.43	21.34	31.55
	(4.30)	(4.10)	(4.30)	(4.23)
Government	13.21	16.19	11.77	13.59
	(5.6)	(4.92)	(4.50)	(3.87)

[1] Data are from the Bureau of Economic Analysis, the Current Population Survey, and authors' calculations. Standard deviations are in parenthesis.

Table 3: Labor Productivity and Density by Industry, 1999

Non-Linear Least Squares and Generalized Method of Moments Instrumental Variables Estimation[1]

$$\ln\left(\frac{Y_s}{N_s}\right) = \alpha + \ln\left(N_s^{-1}\sum_{c\in s} a_c \left(\frac{n_c}{a_c}\right)^{\theta}\right) + \eta \ln edu_s + \varepsilon_s$$

	NLS		IV=Density in 1977		IV = Density in 1994	
	Edu	Density	Edu	Density	Edu	Density
Total GSP	0.19	1.08**	0.30	1.08**	0.27	1.08**
	(0.75)	(0.01)	(0.73)	(0.02)	(0.74)	(0.02)
Agriculture	−0.12	0.95	−0.17	0.91*	−0.14	0.93*
	(0.37)	(0.05)	(0.21)	(0.05)	(0.22)	(0.04)
Construction	1.73**	1.10**	1.68**	1.09**	1.69**	1.09**
	(0.6)	(0.02)	(0.48)	(0.02)	(0.47)	(0.02)
Manufacturing	2.50**	0.99	2.33**	1.01	2.30**	1.01
	(0.99)	(0.03)	(1.10)	(0.03)	(1.07)	(0.03)
TPU	1.05	1.02	1.05	1.03	1.20	1.02
	(1.06)	(0.02)	(0.97)	(0.02)	(0.98)	(0.03)
Wholesale Trade	0.49	1.07**	0.46	1.07**	0.59	1.06**
	(0.71)	(0.01)	(0.56)	(0.02)	(0.62)	(0.02)
Retail Trade	−1.47**	1.07**	−1.53**	1.07**	−1.51**	1.07**
	(0.66)	(0.01)	(0.74)	(0.01)	(0.75)	(0.01)
FIRE	2.30*	1.11**	2.74**	1.10**	2.66**	1.10**
	(1.29)	(0.02)	(1.17)	(0.02)	(1.17)	(0.02)
Services	−1.39*	1.12**	−1.29	1.11**	−1.32	1.11**
	(0.74)	(0.01)	(0.94)	(0.01)	(0.94)	(0.01)
Government	1.53**	1.07**	1.69**	1.07**	1.64**	1.07**
	(0.52)	(0.01)	(0.61)	(0.01)	(0.59)	(0.01)

[1] The instrumental variable used is the lagged employment share weighted average workers per acre at the county level for each state in respective years. Heteroskedasticity-consistent White standard errors appear in parenthesis. A * (**) indicates significance at the 10% (5%) level. Inference regarding θ, the coefficient on the non-linear density term, is relative to the null value of 1.

Table 4: Labor Productivity and Density by Industry, 1982-1999
NLS and GMM-IV Estimation with State and Time Fixed Effects[1]

$$\ln\left(\frac{Y_{st}}{N_{st}}\right) = \alpha + \ln\left(N_{st}^{-1}\sum_{c\in st} a_c\left(\frac{n_{ct}}{a_c}\right)^{\theta}\right) + \eta \ln edu_{st} + \upsilon_s + \zeta_t + \mu_{st}$$

	NLS		IV=Lagged Density	
	Edu	Density	Edu	Density
Total GSP	0.59**	0.99	0.68**	1.05**
	(0.13)	(0.02)	(0.16)	(0.02)
Agriculture	−0.04	0.77**	−0.12	0.14**
	(0.11)	(0.08)	(0.11)	(0.13)
Construction	0.63**	1.25**	0.92**	1.56**
	(0.11)	(0.02)	(0.17)	(0.11)
Manufacturing	0.67**	1.08**	0.74	1.20**
	(0.29)	(0.03)	(0.47)	(0.04)
TPU	0.21	0.84**	0.20	0.78**
	(0.15)	(0.03)	(0.15)	(0.04)
Wholesale Trade	0.13**	0.95**	0.13	0.97
	(0.06)	(0.01)	(0.08)	(0.02)
Retail Trade	−0.16	0.97	−0.16	1.04*
	(0.10)	(0.02)	(0.11)	(0.02)
Fire	0.98**	0.91**	1.01**	1.04
	(0.17)	(0.03)	(0.18)	(0.05)
Services	0.30**	0.94**	0.3**	0.96*
	(0.14)	(0.02)	(0.15)	(0.03)
Government	0.20**	0.96**	0.19**	0.93**
	(0.06)	(0.02)	(0.06)	(0.03)

[1] The instrumental variable used is the employment share weighted average workers per acre at the county level for each state lagged five years. Autoregressive and heteroskedasticity-consistent Newey-West standard errors appear in parenthesis. A * (**) indicates significance at the 10% (5%) level. Inference regarding θ, the coefficient on the non-linear density term, is relative to the null value of 1.

Table 5: Labor Productivity and Density, Total GSP 1982-1999

Non-Linear Least Squares Estimation and GMM Instrumental Variables Estimation with State and Time Fixed Effects[1,2]

$$\ln\left(\frac{Y_{st}}{N_{st}}\right) = \alpha + \ln\left(N_{st}^{-1}\sum_{c\in st} a_c\left(\frac{n_{ct}}{a_c}\right)^{\theta}\right) + \eta \ln edu_{st} + \upsilon_s + \zeta_t + \mu_{st}$$

	NLS		IV=Lagged Density		IV=Birthrate	
	(1)	(2)	(1)	(2)	(1)	(2)
Education	0.59**	0.52**	0.68**	0.75**	0.01	1.07**
	(0.13)	(0.13)	(0.16)	(0.17)	(0.30)	(0.02)
Density	0.99	0.93**	1.05**	1.07**	0.61**	1.08**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.13)	(0.02)
		Wald[3]		Wald		Wald
Lags of Bartik		76.00		28.05		28.06
		(< 0.01)		(< 0.01)		(< 0.01)
Lags of Defense		9.67		28.43		28.43
		(0.08)		(< 0.01)		(< 0.01)
R2	0.94	0.94	0.94	0.93	0.90	0.93
N. Obs	900	864	900	864	897	864

[1] Autocorrelation and heteroskedasticity consistent Newey-West standard errors appear in parenthesis. A * (**) indicates significance at the 10% (5%) level. Inference regarding θ, the coefficient on the non-linear density term, is relative to the null value of 1.

[2] The Wald test is of the null that the coefficients are jointly zero. P-values are in parenthesis.

[3] The Wald test is of the null that the coefficients are jointly zero. P-values are in parenthesis.

References

[1] Bartik, T., 1985, Business location decisions in the United States, Journal of Business and Economic Statistics 3, 14-22.

[2] Bartik, T., 1991, Who benefits from state and local economic development policies? Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.

[3] Barro, R., 1981, Output effects of government purchases, Journal of Political Economy 89(6), 1086-1121.

[4] Baumgardner, J., 1988, The division of labor, local markets, and worker organization, Journal of Political Economy 96(3), 509-527.

[5] Black, D., 1998, Local human capital externalities: educational segregation and inequality, London School of Economics, mimeo.

[6] Blanchard, O. and L. Katz, 1992, Regional evolutions, Brookings Papers on Economic Activity 1992(1), 1-75.

[7] Burchfield, M., Overman, H., Puga, D. and M. Turner, 2006, Causes of sprawl: a portrait from space, Quarterly Journal of Economics 121(2), 587-633.

[8] Ciccone, A., 2002, Agglomeration effects in Europe, European Economic Review 46(2): 213-227.

[9] Ciccone, A. and R. Hall, 1996, Productivity and the density of economic activity, American Economic Review 86, 54-70.

[10] Combes, P., Duranton, G., Gobillon, L., and S. Roux, 2010, Estimating agglomeration effects with history, geology, and worker fixed-effects, in E. Glaeser (ed.), Agglomeration Economics. Chicago, IL: University of Chicago Press, 15-66.

[11] Costello, D., 1993, A cross-country, cross-industry comparison of productivity growth, Journal of Political Economy 101, 207-222.

[12] Costa, D. and M. Kahn, 2000, Power couples: changes in the locational choice of the college education, Quarterly Journal of Economics 115(4), 1287-1315.

[13] Courant, P. and A. Deardorff, 1992, International trade with lumpy countries, Journal of Political Economy 100, 198-210.

[14] Duranton, G. and D. Puga, 1999, Nursery cities: urban diversity, process innovation, and the life-cycle of products, American Economic Review 91(5), 1454-1477.

[15] Ellison, G. and E. Glaeser, 1997, Geographic concentration in U.S. manufacturing industries: a dartboard approach, Journal of Political Economy 105, 889-927.

[16] Ellison, G., Glaeser, E. and W. Kerr, 2010, What causes industry agglomeration? Evidence from coagglomeration patterns, American Economic Review 100(3), 1195-1213.

[17] Foote, C., 2000, Interstate migration in the United States, 1979-2000: a gross flows approach, Harvard University, mimeo.

[18] Foote, C., 2001, Young persons and state-level unemployment: a case of increasing returns? Harvard University, mimeo.

[19] Fujita, M., 1988, A monopolistic competition model of spatial agglomeration: a differentiated product approach, Regional Science and Urban Economics 18(1), 87-124

[20] Fujita, M., 1989, Urban Economic Theory, Land Use and City Size, Cambridge, UK: Cambridge University Press.

[21] Fujita, M., 1998, Monopolistic competition and urban systems, European Economic Review 37, 308-315.

[22] Fujita, M., P. Krugman, and A. Venables, 1999, The spatial economy: cities, regions, and international trade, Cambridge, MA: MIT Press.

[23] Fujita, M. and T. Mori, 1997, Structural stability and evolution of urban systems, Regional Science and Urban Economics 27, 399-422.

[24] Gan, L. and Q. Li, 2004, Efficiency of thin and thick markets, NBER Working Paper No. 10815.

[25] Glaeser, E. and J. Gottlieb, 2009, The wealth of cities: agglomeration economies and spatial equilibrium in the United States, Journal of Economic Literature 47(4), 983-1028.

[26] Glaeser, E., Gyourko, J. and R. Saks, 2006, Urban growth and housing supply, Journal of Economic Geography 6, 71-89.

[27] Glaeser, E. and D. Mare, 2001, Cities and skills, Journal of Labor Economics 19(2), 316-342.

[28] Hanson, G., 1996, Localization economies, vertical organization, and trade, American Economic Review 86, 1266-1278.

[29] Hanson, G., 2005, Market potential, increasing returns, and geographic concentration, Journal of International Economics 67(1), 1-24.

[30] Hanson, G., 2001, Scale economies and the geographic concentration of industry, Journal of Economic Geography 1(3), 255-276.

[31] Harris, T. and Y. Ioannides, 2000, Productivity and metropolitan density, Tufts University, mimeo.

[32] Helsey, R. and W. Strange, 1990, Matching and agglomeration economies in a system of cities, Regional Science and Urban Economics 20(2), 189-212.

[33] Henderson, J., 1974, Optimum city size: the external diseconomy question, Journal of Political Economy 82, 373-388.

[34] Henderson, J., 1986, Efficiency of resource usage and city size, Journal of Urban Economics 19, 47-70.

[35] Henderson, J., 1988, Urban Development: Theory, Fact, and Illusion, Oxford, UK: Oxford University Press.

[36] Henderson J., 2003, Marshall's scale economies, Journal of Urban Economics 53(1), 1-28.

[37] Ho, M., Jorgenson, D., and Stiroh, K., 2002, Is there enough high-tech investment: U.S. capital input growth, Harvard University, mimeo.

[38] Holmes, T., 1999, Localization of industry and vertical disintegration, Review of Economics and Statistics 81(2), 314-325.

[39] Jaeger, D., 1997, Reconciling the old and new census bureau education questions: recommendations for researchers, Journal of Business and Economics Statistics 15, 300-309.

[40] Jaffe, A., Trajenberg, M. and R. Henderson, 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, Quarterly Journal of Economics 108(3), 577-598.

[41] Ke, S., 2009, Agglomeration, productivity, and spatial spillovers across Chinese cities, The Annals of Regional Science 45(1), 157-179.

[42] Kolko, J., 1999, Can I get some service here? Information technology, service industries and the future of cities, SSRN working paper, available at `http://ssrn.com/abstract-985712`.

[43] Krugman, P., 1991, Increasing returns and economic geography, Journal of Political Economy 99, 483-499.

[44] Lucas, R., 1988, The mechanics of economic development, Journal of Monetary Economics 22, 3-42.

[45] Marshall, A., 1920, Principles of Economics, New York: Macmillan.

[46] Overman, H. and D. Puga, 2010, Labor pooling as a source of agglomeration: an empirical investigation, in E. Glaeser (ed.), Agglomeration Economics. Chicago, IL: University of Chicago Press, 133-150.

[47] Rosenthal S. and W. Strange, 2001, The determinants of agglomeration, Journal of Urban Economics 50(2), 191-229.

[48] Rosenthal S. and W. Strange, 2004, Evidence on the nature and sources of agglomeration economies. in V. Henderson and J. Thisse (eds.), Handbook of Regional and Urban Economics, Vol. 4. Amsterdam: North-Holland, 2119-2171.

[49] Saks, R., 2008, Job creation and housing construction: constraints on metropolitan area employment growth, Journal of Urban Economics 64(1): 178-195.

[50] Topel, R., 1986, Local labor markets, Journal of Political Economy 94: s111-s143.

[51] Wheeler, C., 2006, Productivity and the geographic concentration of industry: the role of plant scale, Regional Science and Urban Economics 36(3), 313-330.