PRODUCTIVITY AND METROPOLITAN DENSITY*

Timothy F. Harris
and
Yannis M. Ioannides

Department of Economics
Tufts University
Medford, MA 02155
tharris@alumni.tufts.edu
(617) 627-3294. yioannid@tufts.edu

May 15, 2000

Abstract

This paper evaluates the relationship between urban productivity and density using data on metropolitan areas. This is an alternative measure of the urban economy to the one employed by Ciccone and Hall (1996). They used data on output and education by state and employment and education by county, which excludes agricultural and mining sectors. Instead, our U.S. metropolitan area data are defined contemporaneously for the five available census years from 1950 to 1990. These data allow us to conduct both cross-sectional and panel analyses. Furthermore, since we use a model where income is a linear function of density, these data allow us to evaluate the urban system in its own right. Our results replicate the findings of Ciccone and Hall (1996). We find that a doubling of population density leads to about a 6% increase in productivity. Our results establish an important role for Jacobs externalities, measured by metropolitan area population.

* Research support by the National Science Foundation and by the John D. and Catherine T. MacArthur Foundation is gratefully acknowledged. Anna Hardman, Steve Marrone, and Jeffrey Zabel provided excellent comments.

1 Introduction

The economics of agglomeration is an old question that began with Alfred Marshall (1920), who sought to explain the tendency of firms to locate near each other. He argued that when an industry concentrates the proximity of similar firms encourages technological and informational spillovers. Innovative ideas are rarely kept secret when employees from different firms can easily get together and talk, gossip or even spy on one another, and a high density of economic activity facilitates such an exchange of information. Marshall emphasized the role of industrial clustering in boosting productivity. The importance of urban areas for human creativity has also been emphasized by Jane Jacobs (1969), whose work has had a major effect in popularizing the notion of urbanization externalities, whereby different industries locating near one another confer advantages to all. The most recent in a series of empirical tests of Marshall's localization externalities is a direct examination of the impact of spatial employment density on productivity by Ciccone and Hall (1996).

This paper seeks to replicate the work of Ciccone and Hall by using metropolitan area (MA) data both in a cross-sectional and panel framework. We evaluate the role of population density directly, in contrast to the much more complex employment density index and non-linear estimation that Ciccone and Hall employed. Our data set on MAs combines the five census years from 1950 to 1990. This data set allows us to examine the impact of agglomeration in U.S. cities by means of an alternative measure of the urban economy, namely by focusing our analysis at the metropolitan level rather than at the state or county level.

To anticipate our results, we highlight the three possible relationships between population density and income: (1) there is a significant direct relationship which indicates an income premium and, equivalently, enhanced worker productivity as a result of urban externalities (labor pooling, reduced transportation costs, technological spillovers); (2) there is no significant relationship indicating that any positive effects from agglomeration are offset by negative congestion externalities; (3) there is an inverse relationship indicating that the negative congestion effects of greater density exceed the positive agglomeration effects. Thus, the results will either confirm or deny the theoretical work of Marshall and Jacobs and the empirical work of Ciccone and Hall. The relationship between population size and income will indicate the extent of Jacobs externalities. A positive relationship indicates that Jacobs externalities are present, while an insignificant or negative relationship indicates that size, regardless of the diversity that it brings, does not lead to increasing returns.

We identify which relationships exist using ordinary least squares (OLS) regression that, in principle, mimic the approach of Ciccone and Hall. We also make use of the unique advantages of a local metropolitan data set to evaluate the effectiveness of investigating agglomeration at the state level, as done by Ciccone and Hall. Furthermore, we utilize our data both as cross-sections and as panel data, allowing us to understand how changes in spatial density have affected productivity over time.

The results of our cross-sectional analyses strongly support the findings of Ciccone and Hall (1996) and indicate that a doubling of population density leads to a 6% increase in productivity, measured in terms of income per person. The effect is consistent and highly significant. Although, the panel results are less supportive of the role of density, they help establish a very significant role for population size in determining productivity over time. We obtain similar results when we use wages as an alternative measure of productivity.

The following section presents briefly the theoretical model and empirical approach of Ciccone and Hall. Section 3 presents the data. Section 4 presents the methodology and results. Section 5 concludes.

2.1 Ciccone and Hall's Model

Despite its strong theoretical foundation, an empirical evaluation of the role of density has long been absent from economic literature. Previous studies of agglomeration [Sveikauskas (1975), Segal (1976), Moomaw (1981, 1985), Henderson (1986)] have largely focused on the effects of city population or industry employment on productivity and neglected the role of density. Ciccone and Hall (1996) discuss two different models, one based on local geographic externalities and the other based on a diversity of local intermediate services. They show that these two models are equivalent in that they both predict the same relationship between county employment density and state labor productivity. Specifically, their model may be summarized as follows. Let α be the production elasticity of employment and λ the elasticity of the externality which measures the effect of agglomeration. If q_c, n_c, and a_c denote respectively, output, employment, and area of county c, then output per acre in county c turns out to be a function of employment per acre:

(1) [2] $$\frac{q_c}{a_c} = \left(\frac{n_c}{a_c}\right)^{\gamma},$$

where γ is the elasticity of output at the county level with respect to employment, and the equation number in brackets is as in ibid. Ciccone and Hall show that it is through the product γ, where $\gamma = \alpha\lambda$ and combines the net effect of congestion versus agglomeration, that density affects productivity in their model. If γ exceeds one, agglomeration effects dominate congestion indicating that an increase in the density of economic activity leads to increasing returns to scale and higher productivity. If γ is less than one, then congestion effects dominate agglomeration which indicates that the added congestion from an increase in density actually leads to lower productivity. Solving (1) for average labor productivity, i.e. q_c/n_c, we get $\frac{q_c}{n_c} = \left(\frac{n_c}{a_c}\right)^{\gamma-1}$ where γ-1 is the elasticity of output per worker with respect to employment. By aggregating this to the state level we get

(2) [3] $$\frac{Q_s}{N_s} = \frac{\sum_{c \in C_s} n_c^{\gamma} a_c^{-(\gamma-1)}}{N_s},$$

where C_s denotes the set of counties covering state s, Q_s the output in state s, and N_s its employment. This leads to Ciccone and Hall's density index:

(3) [4] $$D_s(\gamma) = \frac{\sum_{c \in C_s} n_c^{\gamma} a_c^{-(\gamma-1)}}{N_s}.$$

The indirect nature of the evaluation is most apparent from this last equation where we can see that the (γ-1) term representing their density elasticity can only be evaluated indirectly through a larger and more complicated index aggregating county data to the state level. The estimation model, which is consistent with their local geographic externalities model or by the intermediate product variety model, incorporates density index (3) and is given by

(4) [29] $$\log\left(\frac{Q_s}{N_s}\right) = \log\phi + \eta\log h_s + \log D_s(\theta) + u_s,$$

where $\frac{Q_s}{N_s}$ represents productivity and is defined as gross state product less proprietor's income (Q_s) divided by employment (N_s, which is aggregated up from the county level), ϕ is a constant,

η is the elasticity of education, h_s represents education by weighing the workers' years of education by the number of hours worked, u_s is the error term all at state s. It should be noted that θ plays the role of γ and differs from it only in that it is based on efficiency units of labor (labor weighted by experience) instead of raw labor.

Using variations of equation (4) and data on states and counties from 1988, they found that employment density is indeed positively related to labor productivity and thus, by inference, to income. More specifically, they concluded that a doubling of employment density increases average labor productivity by about 6%[1] and, furthermore, that over half of the variance of output per worker across states can be explained by differences in the density of economic activity. This small but significant effect on labor productivity of spatial density of employment is consistent with aggregate increasing returns to scale and indicates that productivity variations can be attributed to differences in economic density.

2.2 An Evaluation of Ciccone and Hall's Model

While Ciccone and Hall's model is an important one, particularly because it is the first of its kind to incorporate density so explicitly, it does not answer all of our questions concerning the role of density. One of the most noticeable omissions from their model is population size. While density rather than size may be a more accurate determinant of productivity, the two are not by any means mutually incompatible. By excluding an explicit population variable, Ciccone and Hall restricted the ability of their model to detect Jacobs externalities which are evident only

[1] If the fact that a 100% increase in density leads to a 6% increase in productivity does not seem economically significant, one must consider that density ranges from a minimum of about 14 to around 11,000 people per square mile in our sample. This allows for significant productivity differentials between cities. For example, given Ciccone and Hall's findings we could conclude that in 1990 New York, NY with a density of 7447.6 is about 36% more productive than New Haven, CT with a density of 1232.8, and about 106% more productive than Louisville, KY with a density of 420.4.

if city size is directly related to productivity. If Jacobs externalities are significant in explaining productivity, which they may be considering their demonstrated role in determining city growth (Glaeser et al, 1992), then we would expect to see higher productivity in larger metropolitan areas that can support a greater diversity of industries. Excluding population from the model, under such circumstances, could have led to an overestimation of the role of density. By estimating both with and without population size, we seek to control for Jacobs externalities in explaining productivity and to get a better estimate of the impact of density.

Another issue that may have affected their results is the decision to perform estimations with state level data. While their index, equation (3), is an ingenious one, the relationship between productivity and density is an urban one, particularly within the context of either Marshall or Jacobs externalities. Estimating this relationship at the state level, when each state contains on average 6.7 metropolitan areas, makes one wonder whether it would still be found at a metro level. Given that urban economic interactions may both transcend, and be more specific than, state divisions, more specific data is needed to evaluate these urban concepts. Ciccone and Hall were certainly aware of this, as they noted that average density of activity for a state was a meaningless concept since most of the area of the U.S. is empty space and holds no economic activity at all. Despite their attempts to compensate by aggregating up county density for their density index, we feel that data gathered directly from the metropolitan areas provide an alternative way to evaluate the role of density. Doing so also allows us to test this assumption by grouping our metropolitan data by state and comparing the grouped (state) results to our original estimations.

Ciccone and Hall's analysis was an indirect one not only because of the type of data they used but also because of how they incorporated density into their model. We use the simplest and most direct measure of density available, population density, and we evaluate income as a direct function of density. Thus, our approach is more direct both because we use more specific and more relevant metropolitan area data and because we estimate income directly as a function of density.

Ciccone and Hall's reliance completely on cross-sectional analyses alone is also a problem in evaluating productivity differentials. Henderson (1997) discusses how such cross-sections can ignore unobserved fixed factors that may underlie productivity differences, such as those arising from the region's history. Henderson refers to these unobserved factors as location fixed effects, examples of which include varying regional resource endowments, cultural influences on legal, political, and institutional arrangements, and skill-specific immobile portions of the labor force. The existence of such unobserved location fixed effects leaves open the possibility that the estimated determinants of productivity in a cross-sectional analysis are biased. Because our data allow us to evaluate the relationship both using repeated cross-sections of the data and as a panel, we sidestep this problem and are better suited to establish whether or not there is a consistent relationship.

Our approach follows Ciccone and Hall's closely but differs from theirs because we allow for population size, estimation at the metropolitan level, and estimation of both the cross-section and panel dimensions of the data. While the economic phenomena explored are the same, the decision to use metro area data requires the use of different variables to represent similar concepts. Our use of wages as a measure of urban productivity is arguably an improvement over gross state product (GSP) less proprietors' income as a measure of labor productivity. We define education is defined as the percent of 16 to 19 year olds who are in school, while they use workers' years of education weighted by the number of hours worked.

Density is population density instead of a density index based on the density of employment. Our estimation model is:

$$\log \pi_{it} = \phi_t + \eta_t \log h_{it} + \theta_t \log D_{it} + \rho_t \log P_{it} + u_{it}, \tag{5}$$

where π_{it} is productivity, h_{it} is education, D_{it} is population density, P_{it} is population size, and u_{it} is the disturbance term, where i indexes metropolitan areas and t=1, ..., 5 indicates the five census years 1950 to 1990. The unknown parameters to be estimated, (ϕ_t, η_t, θ_t, ρ_t) may be time varying.

3 Data

Data on urban variables pose serious issues regarding their definitions. Data could be collected by city proper, reflecting the legal city boundaries. While city data would provide a much more precise picture of localized economic activity, we choose not to use it because it neglects the very real phenomenon of metropolitan integration. Functional cities do not, in a real economic sense, necessarily coincide with their legal boundaries, and MA data reflect this fact.

County data also provide a convenient and more natural alternative to state data for the study of urban externalities, but it, too, is flawed for our purposes. Since this paper is principally concerned with productivity in an urban context, we would consider any non-metropolitan county to be noise. Thus, we would be inclined to drop the majority of counties which support little or no economic activity at all. Yet even if one were to gather data only from counties that met some specified urban criteria, county divisions are too small to get an accurate picture of the larger urban entities. Cities like New York and Chicago have integrated to such an extent with surrounding suburbs and cities that they would best be evaluated as a collection of two, three, or more counties. The use of counties then, while a much better alternative to states or cities as defined by legal boundaries, would require a good deal of work in determining which counties to include and which counties to group together before estimation would make sense.

The general concept of an MA, as intended by the Office of Management and Budget (OMB) and the U.S. Bureau of the Census, is that of a large population nucleus, together with adjacent communities having a high degree of social and economic integration with that core. MAs are comprised of one or more entire counties, with the exception of MAs in New England where, because of the large number of neighboring urban areas, cities and towns are the basic geographical units. Over the years MA names have changed, causing some confusion, but the basic definition has changed little over the years. Originally, they were “standard metropolitan areas” (SMAs) in 1949, then “standard metropolitan statistical areas” (SMSAs) in 1959, then “metropolitan statistical areas” (MSAs) in 1983, and finally “metropolitan areas” (MAs) in 1990. To make matters even more confusing, MA is itself a collective term that includes “metropolitan statistical areas” (MSAs), “primary metropolitan statistical areas” (PMSAs), and “consolidated metropolitan statistical areas” (CMSAs). Despite all the different names, we consider all of these terms to be synonymous with the exception of CMSAs, which are not included in our data set. This is, however, inconsequential for purposes of estimation since every CMSA has a corresponding PMSA[2].

[2] Currently, each newly qualifying MSA must have at least one city with more than 50,000 inhabitants and a total metropolitan area of at least 100,000 (75,000 in New England). The county containing the largest city becomes the “central county,” and adjacent counties are then included in the MSA if 50 percent of their population is in the urbanized area surrounding the largest city. Additional “outlying counties” are included in the MSA if they meet specified requirements of commuting to the central counties and other selected requirements of metropolitan character, such as population density and percent urban. An MSA can become a CMSA if it has a population over one million and separate component areas can be identified within the entire area. The CMSAs component parts are designated PMSAs which, because they are much more comparable to MSAs, are included in our data set instead of CMSAs.

4 Estimation and Results

We begin by discussing the estimation of equation (7) as five repeated cross-sections, one for each of the census years available since 1950. The results of this first series of regressions are shown in Table 1. They indicate a consistent and highly significant role for density in explaining income variations. The elasticity of density ranged from a low of 0.0377 in 1950 to a high of 0.1034 in 1990, averaging 0.05616 over the five periods. On average these results confirm the main finding of Ciccone and Hall that a doubling of density leads to about a 6% increase in income. Additionally, above average elasticities in 1970 and 1990 suggest an upward trend for the density coefficient. This suggests that the role of Marshall and/or Jacobs externalities has not been decreasing and may be increasing.

The results from Table 1 also indicate a significant relationship between population and productivity, although the results are not quite as strong nor consistent as for density. The population coefficient is significant for all of the censuses except 1990, when it is only barely insignificant with a p-value of 0.106. Beginning in 1960, the results indicate a downward trend for the elasticity of population. Such a downward trend is evidence that Jacobs externalities have become less important over time, to the extent that they are only marginally significant today. That the elasticities for density are trending upward while the elasticities for population are trending downward indicates that Marshall’s externalities have become increasingly more important in explaining variations in productivity both on the whole and relative to Jacobs externalities. Over time, however, it is clear that population is a significant variable to include in any model of productivity, and that even for the most recent estimations, leaving it out only risks misspecification.

While this first series of regressions give us a solid starting point and establish the importance both of population and density in explaining productivity differentials, a second series of regressions shows the consequences of excluding population in a model of productivity. Table 2 shows the results of estimating equation (7) by OLS without population. As one would expect, the estimates of the density coefficients are higher for each of the cross-sections when we exclude population, indicating that excluding population from a model evaluating productivity could lead to an overestimation of the true role of density. Had Ciccone and Hall included state population as an explicit element of their model, their estimate of the importance of employment density might have been smaller.

In analyzing these results, however, we realize that there may be individual time-invariant variables not captured by our regressors that are the same for a given state over time but that vary across states themselves. Such state effects could represent differing laws between states or different tax codes or even different cultures. We postulate that the error term in (7) is $u_{it} = \nu_i + \varepsilon_{it}$, where ε_{it} is an error term being both non-systematic and identically and independently distributed and where ν_i is a systematic component which represents all of the specific effects of a set of variables that have been omitted or excluded. Such state-specific effects can be incorporated into our model as either fixed or random.

Econometric precedence provides us with some guidance as to whether to include fixed or random effects [Hsiao (1990) p. 33). Random effects are traditionally considered appropriate when data are randomly sampled from a total population with different samples selected in different time periods. In our case, however, we are trying to assess the differences between a specific series of fixed geographic units, namely states, the entire population of which we have complete and non-random access to over time. Our a priori expectation then would be for fixed effects to be an appropriate means of measuring specific effects across states in our model.

To test this a priori expectation, we run all five cross-sections first with random effects and then with fixed effects, in each case both with and without population. With both models estimated, we can now test to see which one is more appropriate with the Hausman specification test. The results of the fixed effects models are in Tables 3 and 4 along with the corresponding Hausman tests.

The results of the Hausman specification test, listed in Table 5, indicate that the random effects model is misspecified in most cases and at any reasonable significance level, with the exceptions being the 1980 and 1990 cross-section with population and the 1990 cross-section without population. Although these results are slightly ambiguous, they do appear to be biased in favor of the fixed effects specification with only three out of the ten tests indication random effects. Since this conclusion affirms our a priori expectation, we conclude that the fixed effects model is indeed the appropriate one and continue our analysis with reference to Tables 3 and 4 only.

When we specify a model with fixed effects, the relationship between density and productivity over time is stronger and more consistent. The elasticity of density ranges from a low of 0.05215 in 1980 to a high of 0.0749 in 1950 when one includes population in the model; a difference of only 0.02275 compared with a difference of 0.0657 between the high and low estimates in our original model (Table 1). Overall, the density coefficient is larger when we include fixed effects and control for population than when we just control for population, averaging 0.061 compared to 0.056. These results complement Ciccone and Hall's findings based on different data that a doubling of employment density leads to an increase in productivity of around 6%. The consistency of the results for the density elasticity over time do not provide any indication of a trend or increasing relevance of density, implying that any perceived trend in the first series of regressions is probably spurious.

In contrast to those for density, the results for the elasticities of population are contradictory to the first series of regressions. When we allow for fixed effects, the estimates appear to be increasing with time instead of decreasing, rising from total insignificance in 1950 to highly significant in 1990. This would imply that Jacobs externalities have become increasingly important in explaining productivity differentials over time. Also, these population elasticities are both smaller and less consistent over time. The elasticity of population ranges from a low of 0.0004 to a high of 0.0379; a difference of 0.0375 compared with the difference between the high and the low estimates without specific effects of 0.0197. The average elasticity of density of population is only 0.0195 compared with 0.0272 when there are no specific effects. While the role of population may not be as grand as the initial estimates implied, it is still very clear that it is an important factor which should be included in any study of productivity and certainly for more recent analyses.

The last step in our analysis is to estimate the model within a panel framework while holding the coefficients constant over time. We report our results in Table 6, where we have also allowed for census dummies. We again test for random effects or fixed effects. The results of the Hausman test (Table 8) strongly and consistently confirm that the model is correctly specified with fixed effects. That the specific effects are fixed over time is much clearer in this panel regression and eliminates any concern over ambiguity in using the fixed effects model.

Limiting ourselves to the MA fixed effects runs, we establish the need to include population in this model by running equation (8) with fixed effects and both with and without the population variable. The results, a relatively high and significant elasticity of 0.0901, strongly supports the inclusion of the population variable. Thus, the final model we use to evaluate the relationship in a panel format with MA fixed effects and population is regression (23).

The panel results are less than encouraging with regards to the role of density in explaining productivity variations. The density elasticity estimate is only 0.0127 and does not support the Ciccone and Hall's conclusion. Furthermore, at 0.0901 the population elasticity in this panel regression is much higher than what we observed for any of the cross-sectional regressions. Notice that when population is removed from the estimation, the density elasticity is almost twice as large. This indicates that one of the reasons the panel regression does not produce similar results to Ciccone and Hall's findings is the importance of population, and by extension of Jacobs externalities, in understanding productivity. A panel analysis, then, provides evidence that Ciccone and Hall's methodology may have overestimated the impact of density.

Another indication as to why we were unable to confirm Ciccone and Hall's results in a panel framework with MA effects is suggested when we use state specific effects. In this way we can see how our data behaved when we allow for some systematic element that is fixed within states and different between them. These regressions are summarized in Table 7 and again the Hausman statistic and population coefficient indicate that a fixed effects model with population, regression (27), is the most appropriate. At 0.0547, the density elasticity is now about what we would expect to find given Ciccone and Hall's analysis and the population elasticity (0.0261) is much lower than when we group by MA. Thus, the difference between confirming and denying Ciccone and Hall's results in a panel framework lies in whether we believe that the specific effects are at the state or MA level. In fact, there are most certainly both state and MA fixed effects, and this second panel regression reveals the importance of considering both types of specific effects. Since the reality is a combination of the two, these results provide additional evidence that Ciccone and Hall's approach led to an overestimation of the role of density.

Lastly, in order to test for consistency in our results, all of the regressions described above were run again with wages instead of median family income. Wages should behave the same way as median family income, but if the results tell a different story than that described above, it would indicate a misspecification in the model. In fact, the results are remarkably similar such that one could insert "wage" into the above analysis in place of "median family income" without any discrepancy. While this confirms our previous analysis, the regression with wages strengthens the case for a significant effect of metropolitan density on urban productivity.

5 Conclusions

Overall, our results confirm the main conclusion of Ciccone and Hall's analysis that on average, a doubling of employment density leads to about a 6% increase in productivity. If there is anything we would add to that conclusion, it is that 6% may be an upper bound over time. The cross-sectional regressions for each of the five recent census years were on average supportive of the "6%" conclusion, with the estimated effect being slightly larger and more consistent when we allowed for state specific effects. The panel regressions were less conclusive concerning the role of density. When we controlled for state specific effects as opposed to MA specific effects, the results were supportive. When instead we controlled for MA effects, the results were not supportive. Overall, the results point solidly to a relationship between density and productivity which is direct, positive, and highly significant.

From among the models we estimated the one which bests mimics Ciccone and Hall, regression (10), provides an estimate of the density coefficient that is almost twice as large as theirs. Because we reject this in favor of one that provides much lower estimates, we conclude that their approach overestimates the relationship for a given data set. However, according to our argument, the relationship between density and productivity is better represented at the urban level than at the state level, our data set should illustrate a stronger relationship between density and productivity for any given methodology. The combination of an approach that provides a lower estimate with a data set that better captures the relationship delivers the same result.

Concerning how best to specify a model of urban productivity, both the cross-sectional and panel analyses illustrated that population has a direct, consistent, and highly significant relationship to productivity. To the extent that our results provide empirical evidence for the role of Jacobs externalities in explaining urban dynamics, they also illustrate the importance of including an explicit measure of city size as well as density in understanding productivity differentials.

While there are many different types of data available for such studies as these, our results show that the decision as to which type of data to use is not trivial. In the panel regressions, the coefficients for both density and population were significantly different when we controlled for state fixed effects as opposed to MA fixed effects. The theory tells us that both of these effects are present in reality, placing a premium on data that can evaluate both cases. Thus, state data, while superior in certain respects, limits our ability to explain productivity differentials. These results also provide evidence that over time 6% is an upper bound, since the density elasticity was much smaller when we allowed for MA fixed effects.

These findings provide support for the continued empirical study of the role of density in determining productivity. By evaluating and comparing several variations of a parsimonious model that incorporates both Jacobs and Marshall externalities, we have been able to illustrate some important benchmarks for studies of urban productivity, while confirming the benchmark of 6% for the effect of density. Future research should refine the relationship between urban density, size, and productivity, possibly by means of micro data.

REFERENCES

CensusCD [CD-ROM] (1990). GeoLytics, Inc.

Ciccone, Antonio and Hall, Robert E. (1996). "Productivity and the Density of Economic Activity." *American Economic Review*, v86, n1, pp. 54-70.

Dobkins, Linda Harris and Ioannides, Yannis M. (2000). "Dynamic Evolution of the Size Distribution of U.S. Cities," in Jean-Marie Huriot and Jacques-François Thisse, eds., *Economics of Cities*, Cambridge: Cambridge University Press.

Glaeser, Edward L., et al. (1992) "Growth in Cities." *Journal of Political Economy*, v100, n6, pp. 1126-52

Gujarati, Damodar (1995). *Basic Econometrics*. New York, NY: McGraw-Hill, Inc.

Henderson, Vernon (1974). "The Sizes and Types of Cities." *American Economic Review*, v64, n4, pp. 640-56.

Henderson, Vernon (1986). "Efficiency of Resource Usage and City Size." *Journal of Urban Economics*, v19, n1, pp. 47-70.

Henderson, Vernon (1997). "Externalities and industrial development." *Journal of Urban Economics* v42, pp.449-70

Henderson, Vernon (1999). "Marshall's Scale Economies." NBER Working Paper #7358.

Hsiao, Cheng. *Analysis of Panel Data.* New York, NY: Cambridge University Press, 1990.

Jacobs, Jane (1969). *The Economy of Cities*. New York, NY: Random House.

Marshall, Alfred (1920). *Principles of Economics*. London: Macmillan.

Moomaw, Ronald L (1981). "Productivity and City Size: A Critique of the Evidence." *Quarterly Journal of Economics*, v96, n4, pp. 675-88

Moomaw, Ronald L (1985). "Firm Location and City Size: Reduced Productivity Advantages as a Factor in the Decline of Manufacturing in Urban Areas." *Journal of Urban Economics*, v17, n1, pp. 73-89.

Segal, David (1976). "Are there Returns to Scale in City Size?" *Review of Economics and Statistics*, v58 n3, pp. 339-50

Sveikauskas, Leo A (1975). "The Productivity of Cities." *Quarterly Journal of Economics*, v89 n3, pp. 393-413.

United States Department of Commerce. Bureau of the Census. "Census of Population and Housing, 1980 [United States]: Summary Tape File 3C." ICPSR study # 8038.

United States Department of Commerce. Bureau of the Census. "County and City Data Book, 1972." ICPSR study # 61.

United States Department of Commerce. Bureau of the Census. "County and City Data Books: 1952, 1956, 1962, 1967." ICPSR study # 12.

Descriptive Statistics

Descriptive Statistics: Mean

	1950	1960	1970	1980	1990	Mean of the Means
Population	502980.2	532779.8	573736.8	532800.7	575300.7	543519.6
LnPop	12.4631	12.542	12.6041	12.5524	12.6328	12.5589
Density	512.0833	591.1934	539.642	427.252	449.3161	503.8974
LnDensity	5.7973	5.7732	5.7004	5.4735	5.5609	5.6611
MedFmInc	3250.982	5834.269	9667.979	19944.91	35318.44	14803.32
LnMdFmIn	8.0764	8.6603	9.1647	9.8905	10.4535	9.2491
MedFmDef	13489.55	19710.37	24917.47	24204.99	27022.52	21868.98
LnMdFmDe	9.4994	9.8777	10.1114	10.0841	10.1858	9.9517
Education	53.91%	63.23%	74.21%	70.03%	81.13%	74.50%
LnEducation	3.9759	4.14	4.3019	4.2431	4.3937	4.2109
WageDe	2827.049	4107.876	4765.289	3525.577	3838.006	3812.959
LnWageDe	7.9309	8.3092	8.4613	8.1557	8.2365	8.21872

Descriptive Statistics: Standard Deviation

	1950	1960	1970	1980	1990	Mean of the Variances
Population	1194991	1078090	1119553	966353.8	978181.4	1067434
LnPop	0.9313	0.9662	0.9989	0.9801	1.001	0.9755
Density	571.5318	1084.37	998.8855	856.5446	859.8047	874.2273
LnDensity	0.9527	1.0968	1.0335	0.99	0.9734	1.0093
MedFmInc	448.3361	849.3708	1456.215	2862.966	7106.152	2544.608
LnMdFmIn	0.1475	0.1526	0.1575	0.1438	0.1908	0.1584
MedFmDef	1860.316	2869.496	3753.131	3474.473	5436.995	3478.882
LnMdFmDe	0.1475	0.1526	0.1575	0.1438	0.1908	0.1584
Education	7.61%	6.98%	6.65%	7.08%	5.07%	6.68%
LnEducation	0.1595	0.1188	0.0998	0.111	0.0681	0.1114
WageDe	488.9316	625.4315	589.6079	539.6186	711.2569	590.9693
LnWageDe	0.1838	0.1521	0.1303	0.1565	0.1805	0.1606

Descriptive Statistics: Minimum

	1950	1960	1970	1980	1990	Mean of the Minimums
Population	56141	51850	55959	57118	56735	55560.6
LnPop	10.9356	10.8561	10.9324	10.9529	10.9462	10.9246
Density	14	14	19	13.43806	11.5	14.38
LnDensity	2.6391	2.6391	2.9444	2.5981	2.4423	2.6526
MedFmInc	1617	2952	4761	12083	17619	7806.4
LnMdFmIn	7.3883	7.9902	8.4682	9.3996	9.7767	8.6046
MedFmDef	6709.543	9972.973	12270.62	14663.83	13480.49	11419.49
LnMdFmDe	8.8113	9.2076	9.415	9.5931	9.509	9.3072
Education	23.51%	36.92%	41.28%	30.00%	45.47%	35.44%
LnEducation	3.16	3.61	3.72	3.4	3.82	3.542
WageDe	1414	2426	2453	2082	1985	2072
LnWageDe	7.25	7.79	7.81	7.64	7.59	7.616

Descriptive Statistics: Maximum

	1950	1960	1970	1980	1990	Mean of the Maximums
Population	12900000	10700000	11600000	9120346	8863164	10636702
LnPop	16.3767	16.1853	16.2641	16.026	15.9974	16.1699
Density	3645	13572	12851	11993.37	11855.3	10783.33
LnDensity	8.2011	9.5158	9.4612	9.3921	9.3805	8.5901
MedFmInc	4262	8745	15834	31784	69824	26089.8
LnMdFmIn	8.3575	9.0762	9.6699	10.3667	11.1537	9.7248
MedFmDef	17684.65	29543.92	40809.28	38572.82	53423.11	36006.76
LnMdFmDe	9.7805	10.2936	10.6167	10.5603	10.886	10.4274
Education	70.37%	80.32%	86.91%	87.66%	92.73%	83.60%
LnEducatic	4.25	4.39	4.46	4.47	4.53	4.42
Wage	4163	7311	6632	5782	6949	6167.4
LnWage	8.33	8.9	8.8	8.65	8.85	8.706

	Table 1 - Crossections of equation (7) Dependent Variable: Ln Median Family Income				
	(1)	(2)	(3)	(4)	(5)
	1950	1960	1970	1980	1990
LnPopDens	0.0377 (0.0135) 1%*	0.038 (0.01) 1%	0.0631 (0.009) 1%	0.0386 (0.0088) 1%	0.1034 (0.0105) 1%
LnPop	0.0317 (0.0133) 5%	0.0362 (0.0112) 1%	0.0265 (0.0092) 1%	0.0252 (0.0087) 1%	0.0165 (0.0102) not
LnEduc	0.2936 (0.0635) 1%	0.3832 (0.0758) 1%	0.5216 (0.0789) 1%	0.3827 (0.0641) 1%	0.6509 (0.1202) 1%
State Random Effects	NO	NO	NO	NO	NO
Adj Rsq	0.2627	0.2771	0.4053	0.2565	0.3998
N	162	210	242	317	323

* Significance levels are indicated as 1%, 5%, 10%, or not

	Table 2 - Crossections of (7), without population Dependent Variable: Ln Median Family Income				
	(6)	(7)	(8)	(9)	(10)
	1950	1960	1970	1980	1990
LnPopDens	0.057 (0.011) 1%	0.0559 (0.0085) 1%	0.0765 (0.0079) 1%	0.0532 (0.0073) 1%	0.1137 (0.0084) 1%
LnPop					
LnEduc	0.2841 (0.0643) 1%	0.3829 (0.0775) 1%	0.5196 (0.08) 1%	0.3655 (0.0645) 1%	0.6243 (0.1194) 1%
State Random Effects	NO	NO	NO	NO	NO
Adj Rsq	0.2410	0.2443	0.3872	0.2392	0.3968
N	162	210	242	317	323

	Table 3 - Crossections of equation (7) Dependent Variable: Ln Median Family Income				
	(11)	(12)	(13)	(14)	(15)
	1950	1960	1970	1980	1990
LnPopDens	0.0749 (0.0163) 1%	0.0632 (0.0124) 1%	0.0543 (0.0128) 1%	0.05215 (0.0114) 1%	0.0604 (0.0132) 1%
LnPop	0.0004 (0.0137) not	0.0124 (0.0115) not	0.0232 (0.0105) 5%	0.0235 (0.0089) 1%	0.0379 (0.0103) 1%
LnEduc	0.0737 (0.0578) not	0.137 (0.0667) 5%	0.27 (0.0735) 1%	0.2717 (0.0562) 1%	0.5011 (0.1118) 1%
State Random Effects	NO	NO	NO	NO	NO
State Fixed Effects	YES	YES	YES	YES	YES
Rsq within Rsq between Rsq overall	0.3365 0.0809 0.1955	0.3214 0.0567 0.2283	0.3326 0.3640 0.3985	0.3174 0.0216 0.2523	0.3384 0.4213 0.3854
Observations Groups	162 43	210 47	242 48	317 51	323 51

	Table 4 - Crossections of (7), without population Dependent Variable: Ln Median Family Income				
	(16)	(17)	(18)	(19)	(20)
	1950	1960	1970	1980	1990
LnPopDens	0.0753 (0.01) 1%	0.0728 (0.0087) 1%	0.0751 (0.0088) 1%	0.0741 (0.0079) 1%	0.0953 (0.0095) 1%
LnPop					
LnEduc	0.0736 (0.0574) not	0.1358 (0.0667) 5%	0.2748 (0.0742) 1%	0.2674 (0.0568) 1%	0.4736 (0.1141) 1%
State Random Effects	NO	NO	NO	NO	NO
State Fixed Effects	YES	YES	YES	YES	YES
Rsq within Rsq between Rsq overall	0.3365 0.0801 0.1948	0.3164 0.0400 0.2072	0.3154 0.2902 0.3733	0.2994 0.0159 0.2306	0.3048 0.4097 0.4001
Observations Groups	162 43	210 47	242 48	317 51	323 51

Table 5 - Corresponding Hausman Specification Test to Tables 3 and 4

(11)	chi2(3) = 25.44;	Prob>chi2 = 0.0000	(16)	chi2(2) = 12.83;	Prob>chi2 = 0.0016
(12)	chi2(3) = 16.32;	Prob>chi2 = 0.0010	(17)	chi2(2) = 15.39;	Prob>chi2 = 0.0005
(13)	chi2(3) = 18.77;	Prob>chi2 = 0.0003	(18)	chi2(2) = 14.03;	Prob>chi2 = 0.0009
(14)	chi2(3) = 4.83;	Prob>chi2 = 0.1846	(19)	chi2(2) = 9.32;	Prob>chi2 = 0.0095
(15)	chi2(3) = 3.49;	Prob>chi2 = 0.3224	(20)	chi2(2) = 1.14;	Prob>chi2 = 0.5669

	Table 6 - Panel of (8) Deflated nationally by CPI-U Dependent Variable: Ln Med Fam Inc			
	(21)	(22)	(23)	(24)
	Random Effects	Random Effects	Fixed Effects	Fixed Effects
LnPopDens	0.033 (0.0057) 1%	0.0546 (0.0052) 1%	0.0127 (0.0074) 10%	0.0226 (0.0076) 1%
LnPop	0.0561 (0.0066) 1%		0.0901 (0.0105) 1%	
LnEduc	0.0483 (0.0294) not	0.0363 (0.0303) not	-0.0496 (0.031) not	-0.075 (0.032) 5%
MA Random Effects	YES	YES	NO	NO
MA Fixed Effects	NO	NO	YES	YES
Rsq within Rsq between Rsq overall	0.9380 0.3431 0.7284	0.9328 0.3424 0.7272	0.9399 0.2262 0.6840	0.9350 0.2026 0.6767
Observations Groups	1254 334	1254 334	1254 334	1254 334
Time Dummies	YES	YES	YES	YES

	Table 7 - Panel of (8) Deflated nationally by CPI-U Dependent Variable: Ln Med Fam Inc			
	(25)	(26)	(27)	(28)
	Random Effects	Random Effects	Fixed Effects	Fixed Effects
LnPopDens	0.0525 (0.0054) 1%	0.0747 (0.004) 1%	0.0547 (0.0056) 1%	0.0766 (0.0041) 1%
LnPop	0.0271 (0.0046) 1%		0.0261 (0.0047) 1%	
LnEduc	0.1985 (0.03) 1%	0.1921 (0.0303) 1%	0.1913 (0.03) 1%	0.1863 (0.0304) 1%
State Random Effects	YES	YES	NO	NO
State Fixed Effects	NO	NO	YES	YES
Rsq within Rsq between Rsq overall	0.8284 0.3581 0.7529	0.8240 0.2800 0.7477	0.8284 0.3466 0.7527	0.8240 0.2715 0.7473
Observations Groups	1254 51	1254 51	1254 51	1254 51
Time Dummies	YES	YES	YES	YES

Table 8 - Corresponding Hausman Specification Test to Tables 6 and 7

(23) chi2(7) = 162.33; Prob>chi2 = 0.0000
(24) chi2(7) = 139.65; Prob>chi2 = 0.0000

(27) chi2(7) = 18.93; Prob>chi2 = 0.0084
(28) chi2(7) = 12.77; Prob>chi2 = 0.0469