THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

TOWARD A THEORY OF CITIES: EMPIRICAL EVIDENCE OF AND A PROPOSED EXPLANATION FOR URBAN SCALING

DANE R. WILBURNE
Spring 2012

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Economics
with honors in Economics

Reviewed and approved* by the following:

Dr. N. Edward Coulson
Professor of Economics
Thesis Supervisor

Dr. James R. Tybout
Professor of Economics
Thesis Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Are cities, in some sense, scaled versions of one another? Empirical evidence presented herein suggests an affirmative answer to this question. I propose an explanation for this phenomenon in terms of simple network theory. The basic premise is that cities consist of two types of networks: a social network and an infrastructural network. The fundamental differences in the structure of these networks can account for the divergent behavior of various magnitudes- so-called urban quantities- associated to cities.

Table of Contents

I. Introduction

American cities are like magnets exerting an attractive force on the country's iron filings of citizens. In 2010, nearly eighty-four percent of Americans lived in the United States' three hundred sixty-six metropolitan statistical areas (U.S. Census Bureau [Census], 2011). Between 2000 and 2010, total U.S. population increased by over twenty-seven million people. Over twenty-five million of those new residents live not in rural America, but rather in dense urban centers (Census, 2011). Over the same ten-year span, the population of some of America's largest metropolises, such as Dallas-Fort Worth, Texas, Houston, Texas, and Atlanta, Georgia, swelled by more than twenty-three percent (Census, 2011). The trend is clear: every day, more and more Americans are choosing to live in cities.

Increasing urbanization is not, however, a uniquely American phenomenon. Harvard economist Edward Glaeser notes that "five million more people every month live in the cities of the developing world, and in 2011, more than half the world's population is urban (Glaeser, 2011, p. 1)." Furthermore, "the number of urban areas with over one million people is expected to grow by over forty percent between 2000 and 2015 (Crane & Kinzig, 2005, p. 1225)." So not only are people all over the world moving to existing cities, but new cities are springing up across the globe.

Given the worldwide trend of increasing urbanization, it is reasonable to ask how changes in urban population impact other characteristics of cities. In this paper, I will consider a set of indicators, or *urban quantities*, associated to cities and examine how they are affected by population changes. More specifically, I will draw on the work of Bettencourt, Lobo, Helbing, Kühnert, and West (2007) in which they propose the existence of power law scaling relationships for a wide range of urban quantities. I have collected and analyzed yearly data from American cities over the period 2001 to 2010 and tested for the presence of power law scaling relationships between urban quantities and the size of a city as measured by its population.

If urban quantities can be described by power laws, it would suggest that cities scale with surprising regularity. The fundamental question then becomes, what are the forces or processes underlying the growth of cities that give rise to scaling? In other words, why do cities with vastly different people, firms, histories, and geographies follow such similar patterns of growth? The answer to these questions, as has been proposed in Bettencourt et al. (2007), is that many urban quantities are subject to the properties of networks. It is my aim here to provide further evidence for this hypothesis and to provide new insights as to why it may be true. Toward that end, I will describe cities in terms of networks and offer some simple applications of network theory. I think that this approach is fruitful for explaining the observed scaling relationships.

Section II provides a brief overview of the relevant literature and introduces terms and concepts that are crucial to the main part of the paper. In Section III, I explain power law scaling, the fundamentals of network theory, and summarize the results of the Bettencourt et al. paper (2007). Section IV contains a discussion of my data collection methods and sources, as well as a description of the econometric techniques used to analyze the data. The results and implications of my analysis are presented in Section V and concluding remarks are offered in Section VI.

II. Literature Review and Key Terms and Concepts

As with many topics in economics, the origins of urban economics can be traced back to Adam Smith. In his classic *The Wealth of Nations*, Smith notes that commercial production tends to cluster in specific locations rather than spread uniformly across space.

> Smiths, carpenters, wheel-wrights, and plough-wrights, masons, and bricklayers, tanners, shoemakers, and taylors, are people, whose service the farmer has frequent occasion for. Such artiticers too stand, occasionally, in need of the assistance of one another; and as their residence is not, like that of the former, necessarily tied down to a precise spot, they naturally settle in the neighborhood of one another, and thus form a small town or village. The butcher, the brewer, and the baker soon join them, together with many other artificers and retailers, necessary or useful for supplying their occasional wants, and who contribute still further to augment the town. (Smith, 1937, p. 358)

Smith astutely attributes this clustering to the conveniences of proximity and to the existence of transportation costs (Smith, 1937). Modern urban economists use the term *agglomeration economies*- discussed in greater detail below- to describe the cost savings that arise when firms choose to concentrate in a specific geographic location.

Smith's "four stages" theory of sociohistorical development together with his theory of economic growth and his conception of an "urban hierarchy" combine to give an accurate description of many urban phenomena. The four stages theory holds that "a developing society passes through four stages in its evolution from a primitive to a modern state (Stull, 1986, p. 292)." The four stages are: hunting, pasturage, agriculture and commerce (Stull, 1986).

In this framework, cities begin to form in the late agriculture and early commerce stages. Smith states that "agriculture, even in its rudest and lowest state, supposes a settlement, some sort of fixed habitation, which cannot be abandoned without great loss (Smith, 1937, p. 49)." The transition from agricultural settlements to cities begins when a place, due to whatever advantages it may have, is able to produce a surplus of food. The surplus allows some people to specialize in trades other than farming, or, as Smith would put it, the surplus allows for an increase in the division of labor (Smith, 1937).

The division of labor is the key concept in Smith's theory of economic growth. Increases in the division of labor, Smith says, increase productivity and efficiency and thereby lead to a rise in income. It follows that increases in wealth are associated with increased demand for goods and services and population growth, which he terms "extensions of the market (Smith, 1937, p. 356)." Furthermore, increased demand and a larger market lead prices to rise in the short-run, which provides an incentive for producers to increase their output, spurring greater competition in the market. Increased levels of competition lead producers to develop new technologies and further increase the division of labor in an effort to reduce costs (Smith, 1937). Thus, "each increase in the division of labor produces a corresponding increase in income and population and, hence, an extension of the market. The latter in turn feeds back and leads to a further increase in the division of labor, causing the cycle to repeat itself (Stull, 1986, p. 296)." Smith's growth mechanism leads to the conclusion that "per capita income, quality of goods produced, trade, population, and the urban share of population increase together in a mutually reinforcing upward spiral (Stull, 1986, p. 304)."

The final component of Smith's urban economics is his idea of an urban hierarchy, which features a dichotomy between government and commercial centers. For the purposes of this paper, we can ignore Smith's discussion of government centers and focus on his hierarchy of commercial cities. The four-tiered hierarchy of commercial cities consists of the following levels, listed in descending order in terms of size: capitals, great towns, small market towns and country villages, and very small villages (Smith, 1937.) Smith describes in great detail the differences between places in each tier, but from the point of view of this paper, we need only to note that Smith ultimately observes "that many economic magnitudes vary systematically with city size (Stull, 1986, p. 304)." In other words, as we move up the hierarchy, we can expect certain characteristics of cities to change in a predictable way. Specifically, Smith (1937) notes that firm size and profit rates, the cost of living, and wage rates are among the economic magnitudes that exhibit this phenomenon. The goal of this paper is to examine and expand upon recent attempts that aim to study, quantify, and explain the type of systematic variation that Smith observed so many years ago.

A major concern of modern urban economics is the study of agglomeration economies, equivalently known as external economies of scale. As noted above, agglomeration economies describe the situation in which long run average costs decrease as the result of an increase in city size or the size of an industry within a city. To distinguish between a decrease in long run average costs due to the overall size of a city and a decrease in long run average costs due to the size of an industry within a city, we use the term “urbanization economy” to describe the former situation and the term “localization economy” to describe the latter (Rosenthal & Strange, 2006).

How do we account for the existence of agglomeration economies? Alfred Marshall pinpointed three forces that help to explain how they arise: input sharing, labor market pooling, and knowledge spillovers (Rosenthal & Strange, 2006). Input sharing refers to the ability of firms to cheaply purchase goods- which can then be used as inputs to production- manufactured by nearby firms. In other words, firms have access to inexpensive inputs to production due to their proximity to other firms. For firms seeking to hire new employees, the fact that other firms in the same industry are located nearby means that there is likely to be an ample supply of workers who have already acquired skills specific to that industry. This is known as labor market pooling, and it also works in the other direction. Skilled workers are likely to be able to find employment in their industry without moving to another city. Knowledge spillovers occur when information is spread by way of the chance encounters that occur countless times every day. For instance, Penn State University looked to take advantage of knowledge spillovers when it broke ground on The Millenium Science Complex, a state of the art research facility that opened in August, 2011. The Complex houses laboratories and resources for researchers in five separate disciplines: chemistry, engineering, biology, physics, and medicine. Spokesman Paul Ruskin explained that the building was designed so that "diverging cultures… [can] hav[e] coffee together and mov[e] ideas forward. This building is all about cross pollination (Stagliano, 2011, p. A1)." The interactions of smart people coming from different backgrounds result in knowledge spillovers that spur new ideas and enhance productivity in cities just as they do in The Millennium Science Complex. The concept of agglomeration economies will be important when we consider power law scaling relationships below.

One last definition to consider is that of *urban quantities*. I will use the term to describe a broad class of social, economic, and infrastructural indicators associated to cities. Any magnitude related to wealth, poverty, crime, disease, productivity, social activity, or material resources that can be measured and aggregated at the city level can be considered an urban quantity. Examples of some familiar urban quantities include gross metropolitan product, total income from wages, and total employment. Some urban quantities that are perhaps less familiar include the total number of patents originating in a city, the number of new AIDS cases, and the total length of all roads.

III. Power Law Scaling and Networks

Power law scaling relationships permeate biology, economics, finance, and other disciplines. Noted urban economist Xavier Gabaix describes the importance of power scaling laws as follows:

> Despite or perhaps because of their simplicity, scaling questions continue to be very fecund in generating empirical regularities, and those regularities are sometimes amongst the most surprising in the social sciences. These regularities in turn motivate theories to explain them, which sometimes require new ways to look at economics issues. (Gabaix, 2009, 258)

It is with this sentiment in mind that I seek to confirm the existence of scaling relationships for urban quantities.

The general functional form for power scaling laws is given by:

$$f(x) = Cx^{\beta}.$$

Following Bettencourt et al. (2007), I will take $x = N(t)$ to be the population of a city at time t and set $f(x) = Y$, where Y is any urban quantity. This gives the following expression:

$$Y(t) = CN(t)^{\beta}.$$

After taking logarithms, we arrive at the model that we wish to estimate:

$$\log Y = \log C + \beta \log N$$

The exponent β is the main parameter of interest, for it provides insight into how changes in population affect the urban quantity in question.

In Bettencourt et al. (2007), ample evidence is found for the existence of power law scaling relationships in data collected for a wide array of urban quantities. In their words:

> We find robust and commensurate scaling exponents across different nations, economic systems, levels of development, and recent time periods for a wide variety of indicators. This finding implies that, in terms of these quantities, cities that are quite superficially different in form and location, for example, are in fact, on the average, scaled versions of one another in a very specific but universal fashion[.] (Bettencourt et. al, 2007, p. 7303)

Their analysis reveals a clustering of β estimates around three distinct values, indicating that urban quantities can scale in three ways. Bettencourt et al. say that an urban quantity scales superlinearly if $\beta > 1$, sublinearly if $\beta < 1$, and linearly if $\beta \approx 1$. Superlinear urban quantities cluster around $\beta \approx 1.1$-1.3, while sublinear urban quantities cluster at $\beta \approx .8$ and linear quantities have scaling exponents that are not significantly different from one (Bettencourt et al, 2007).

Bettencourt et. al describe this clustering as a "taxonomic universality (2007, p. 7303)" that allows urban quantities to be categorized into three distinct groups. They suggest that urban quantities for which $\beta \approx 1.1$-1.3 are quantities associated with wealth creation and innovation (2007). Examples of this type of urban quantity include GDP, total income from wages, and the number of patents originating in a particular city. They conclude that there are increasing returns to scale associated with these urban quantities, so that a doubling of the population results in a more than doubling of quantities that scale superlinearly (2007). Urban quantities that cluster around $\beta \approx .8$ are associated with infrastructure. Examples include the total area of paved surfaces, the total length of all electrical cables, and the total number of gas stations in a city. Thus, these sublinear quantities exhibit economies of scale (2007). In other words, doubling the population of the city implies a less than doubling of urban quantities that scale sublinearly. Linear urban quantities, those whose β's are not found to be significantly different from one, tend to be associated with individual needs. Examples of this type of urban quantity include total employment and the number of housing units in a city (2007).

Finally, before turning to the results, we will consider some topics in network theory. The analysis in this paper hinges on the observation that cities, in some fundamental sense, can be described as networks. I will follow the standard mathematical conventions for networks adopted by Erdös and Rényi (1960). Namely, a *network* (or *graph*) Γ is an entity that consists of *nodes* (or *vertices*) and *edges*. Nodes can be represented visually as dots and edges are the lines connecting the dots. We will describe later two networks associated with cities-a social network and an infrastructural network- that play fundamentally important roles in urban scaling.

There are several important mathematical properties of networks that are important for this paper. First, given a network Γ with n nodes, there are $\binom{n}{2}$ possible edges. It is not necessary that Γ consist of all possible edges. Thus, if Γ has N edges, we must have $N \leq \binom{n}{2}$. It is helpful to think of $\binom{n}{2}$ as an upper bound for N. A crucial observation to make here is that as n grows, the upper bound for N grows much more rapidly. If l new nodes are added to the network so that the total number of nodes becomes $n + l$, then the new upper bound for N is $\binom{n+l}{2}$. In terms of cities, this means that as population grows, the number of possible interactions between people grows at a faster rate than the rate of population growth.

The notation $\Gamma_{n,N}$ means that Γ is a graph consisting of n nodes and N edges. For each node i of Γ, the *degree* (or *connectivity*) of i is defined to be the number of edges originating at i and is denoted by k_i. Nodes of high degree are said to be *highly connected* or to have *high connectivity*. We denote by $P(k)$ the probability that a node is connected to k other nodes in the network. A network is said to be *scale free* if $P(k)$ follows a power law distribution (Albert and Barabási, 1999). The term scale free refers to the fact that such networks are self-similar at all scales (Albert and Barabási, 1999). The conjecture offered in this paper is that the social networks of cities are scale free, or in other words that cities are self-similar to one another in terms of urban quantities.

Albert and Barabási (1999) introduced a generative model that explains how scale free networks arise. The two key components of the model are *network growth* and the concept of *preferential attachment*. Consider a network $\Gamma_{n,N}$ at initial time t_0. At time t_1, one new node of degree m is added to the network. The model assumes that the probability that a new node will be connected to an existing node i is given by $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$. Stated in words, this means that the probability that a new node is connected to an existing node “depends on the connectivity” of the existing node (Albert and Barabási, 1999, p. 5). At each successive time period, another node is added and attaches to the network according to $\Pi(k_i)$. Albert and Barabási's simulations of this model lead them to conclude that "despite its continuous growth, the system organizes itself into scale-free networks" and that "growth and preferential attachment play an important role in network development (1999, p. 5-6)."

For this paper, the technical details of Albert and Barabási's model are less important than the general lessons it offers. First, we note that their approach offers an explanation as to how scale free networks can arise in a natural way. Furthermore, we observe that their model rests on only two assumptions: that the network grows and that as it grows, new nodes are more likely to be attached to already existing highly connected nodes. The thrust of this paper will be to show how this model, as applied to cities, can explain the observed empirical regularities of urban quantities. To my knowledge, such an application of this model has not surfaced anywhere in the literature.

IV. The Data

Bettencourt et al. (2007) present a compelling case for the existence of power laws and scaling for urban quantities. While their work includes some results on American cities, a great deal of their data is from European and Asian cities. This paper focuses only on verifying the existence of power laws for a large number of American cities for the years 2001 to 2010, whereas the most recent data in Bettencourt et al. (2007) is from 2003.

The political boundaries that circumscribe cities are often too restrictive for the purpose of studying them from a statistical perspective because they fail to include a great number of people and firms that are a part of the urban system. In order to properly describe the relationship between urban quantities and population in an entire urban system, most of the data used in this paper are aggregated at the level of metropolitan statistical area. The only exception is data on the total length of roadways in a given city. Roadway data is for Federal-Aid Urbanized Areas which are "areas with 50,000 or more persons that at a minimum encompasses the land area delineated as the urbanized area by the Bureau of the Census (Federal Highway Administration)." Generally, FAUAs and MSAs coincide so that the analysis is the same when testing for scaling relationships. Both designations are intended to capture the essence of unified urban systems by ignoring arbitrary political boundaries and including outlying populations that are very much a part of a city's makeup. Members of such outlying populations often work within official city limits. This implies that they consume goods and services in the city, utilize the city's infrastructure, and frequently interact with its official residents. Excluding these people from this analysis because they reside on the wrong side of a political boundary would thus be a mistake.

The data for this paper were collected from a wide variety of sources. Population data, unless otherwise noted, and data on the number of firms in a city are from the U.S. Census Bureau. Data for gross metropolitan product, total income from wages, and total employment are from the U.S. Bureau of Economic Analysis. Patent data are from The U.S. Patent and Trademark Office, a division of the U.S. Department of Commerce. Data for violent and property crime are from the Crime in the United States data series published yearly by the Federal Bureau of Investigations. Population data are included in the F.B.I's crime data sets and thus these figures were used in lieu of U.S. Census Bureau population data in regressions involving crime statistics. Data on the length of roadways are from the Highway Statistics Series published by the Federal Highway Administration, a division of the U.S. Department of Transportation. Gas station data is from the 2007 U.S. Economic Census. Data on the number of new AIDS cases in a given city is from the U.S. Center of Disease Control, Divisions of AIDS/HIV Prevention.

To test for scaling relationships, urban quantities and population data were log-transformed and regressed using the method of ordinary least squares, with corrections for heteroskedasticity. In cases where data are across multiple years, observations are pooled and regressed together. Scatter plots are presented in the Appendix.

V. Results and Discussion

All empirical results are consistent with Bettencourt et al. (2007). The urban quantities I tested were all found to scale with city size, with the value of scaling exponent contingent on whether the quantity was related to wealth creation and innovation, infrastructure, or individual needs. I offer explanations below as to why each urban quantity analyzed in this paper exhibits its observed behavior.

Urban Quantity (Y)	β	Std. Error	95% C.I.	Adj-R^2	Obs.	Years
Gross metro. Product	1.1208	0.0036	[1.1138, 1.1277]	0.9607	3660	2001-2010
Tot al income from wages	1.1137	0.0037	[1.1065, 1.1209]	0.9629	3294	2001-2009
Violent crime	1.1463	0.0096	[1.1274, 1.1652]	0.8624	1781	2005-2010
Patents	1.306	0.0246	[1.2581, 1.3547]	0.6383	1762	2006-2010
New AIDS cases	1.2944	0.0555	[1.1848, 1.4041]	0.7532	201	2007-2009

Table 1. Scaling exponents for urban quantities that scale superlinearly

First, let us consider urban quantities that were found to scale superlinearly: gross metropolitan product, total income from wages, violent crime, patents, and the number of new AIDS cases in a city. Recall that superlinear scaling refers to the case where the scaling exponent is greater than one. Table 1 summarizes these results. The scaling exponents in the table are all approximately in the range 1.1-1.3, which means that, for cities in the data set, doubling the population of the city implies a more than doubling of these urban quantities. Thus, bigger cities have more GMP per capita, higher wages per capita, more violent crime per capita and so on.

Consider the State College, Pennsylvania and Roanoke, Virginia metropolitan statistical areas. In 2010, Roanoke's population of 308,780 was almost exactly double that of State College, which had 154,127 residents (U.S. Census Bureau). In the same year, State College's gross metropolitan product was $5,397, while GMP in Roanoke was $11,854 (both figures in millions of chained 2005 dollars) (Bureau of Economic Analysis). So despite being twice the size of State College in terms of population, GMP in Roanoke was ten percent more than twice the GMP of State College.

Gross metropolitan product and total income from wages are urban quantities related to wealth creation. It is no surprise that they scale with city size in this way. In fact, this is exactly the relationship predicted by Adam Smith's theory of city growth and his concept of an urban hierarchy, both of which were described above. We saw that Smith wrote of systematic variation of economic magnitudes with regard to city size, and he outlined a theory that explained why we should expect such systematic variation. But while violent crime, the number of new AIDS cases, and the number of new patents originating in a city meet the definition of urban quantity that I adopted for this paper, they are not generally considered to be "economic magnitudes." Thus, the question becomes, why do the number of violent crimes, the number of new AIDS cases, the number of patents, GMP, and total income from wages all scale in the same way with respect to city size? Why do we observe roughly the same scaling exponent in all five cases?

In a broad sense, these five urban quantities are all the product of social interactions. Gross metropolitan product- the total value of goods and services produced in a city- measures the total output of all the firms in a city, and firms are collections of individuals. Wages are paid to individuals by firms. Violent crime occurs when an individual assaults, rapes, murders, or otherwise physically harms another person. AIDS is transmitted when a carrier comes into sexual contact with another individual. A large portion of patent applications are filed by multiple authors who are working on research and development teams for their firms. Thus, there is a clear social component to these five urban quantities, and it is this social component that drives the increasing returns to scale that we observe.

How does the social aspect of these urban quantities manifest in increasing returns to scale? Consider gross metropolitan product. There are two possibilities: either larger cities have proportionally more firms than smaller cities, or larger cities have proportionally the same amount of firms as smaller cities, but the firms in the larger cities are more productive. Regressions on the number of firms in cities, full results of which are included in the Appendix, reveal a scaling exponent of $\beta \approx 1$. Thus, the number of firms scales linearly as a function of population size. Since larger cities do not seem to have proportionally more firms than smaller cities, this simple analysis suggests that firms in larger cities must be more productive in order for the larger cities to achieve a proportionally larger GMP.

We can account for the higher levels of productivity of firms in larger cities by the existence of agglomeration economies. Agglomeration economies, as noted above, refer to the advantages that firms gain by locating in densely populated areas. Being close to other people and firms allows firms to benefit from knowledge spillovers and labor market pooling, which are both social phenomena. Knowledge spillovers occur when ideas are spread through everyday encounters between people. Labor market pooling refers to the ability of firms to choose from a large supply of people in a city that are equipped with a particular skill set. Thus, proximity to more people, or, in other words, increased social interactions, allows firms to be more productive. In this way, the theory of networks complements nicely the standard economic theory of agglomeration economies.

The observation that firms in larger cities are more productive is widely supported in the literature. Leo Sveikauskas (1975) studied the relationship between measures of productivity and city size as measured by population and concludes that “there are very substantial productivity advantages to urban production (p. 411).” Segal (1976) and Moomaw (1981) report similar findings.

Higher productivity for firms in larger cities can, in turn, help to explain why another urban quantity- total income from wages- scales superlinearly with city size. More productive firms can afford to offer higher wages. Sveikauskas (1975) supports this conclusion: “The evidence indicates that these high wages are feasible essentially because productivity is higher in larger cities (p.410).” Furthermore, labor market pooling allows firms to better match their needs with the skills of potential employees. More qualified employees will tend to earn higher wages. It is thus clear that the social aspects of these two urban quantities- GMP and total income from wages- can help to explain the fact that they scale superlinearly.

Similar reasoning applies to the number of new patents originating in a city. Patents are awarded for “novel, non-obvious” inventions (U.S. Department of Commerce). The flash of insight that leads to such an invention can be spurred by knowledge spillovers, which, as we have seen, occur in the course of everyday interactions. More everyday interactions, and thus more knowledge spillovers, occur where there are more people, but the number of possible interactions between people increases dramatically as the population increases. Notice that this statement is suggested exactly by applying the theory of networks to cities. The upper bound for interactions (number of edges) increases much faster than the population (number of nodes) increases. For further evidence on the social nature of ideas, invention, and innovation, see Johnson (2010).

To make this point more concrete, imagine two people in a room. There is one possible interaction: person one interactions with person two. If a third person enters the room, there are three possible interactions. Persons one and two can interact, persons two and three can interact, and persons one and three can interact. For four people, the number of possible interactions is six, while for five people, it is ten. In general, the number of possible interactions between n people is given by $\binom{n}{2} = \frac{n(n-1)}{2}$. Thus, for a hypothetical city of population one hundred, there are 4,950 possible interactions between two people. For a city with a population of two hundred, the number of possible interactions between two people skyrockets to 19,900. In this example, doubling the number of people increased the number of possible interactions by a factor of four, and this calculation does not even include the number of possible interactions between more than two people. In general, we see that the upper bound for the number of possible interactions in a social network grows faster than the rate of population growth. This analysis provides insight into how knowledge spillovers can lead to increasing returns in the number of patents originating in a city.

The same logic applies to the final two superlinear urban quantities analyzed in this paper, violent crime and the number of new AIDS cases in a city. Violent crimes and transmission of the HIV virus that causes AIDS both require interactions between people. As we have just seen, the number of possible interactions increases rapidly as population grows. It therefore makes sense that we observe scaling exponents greater than one for both of these urban quantities.

The key take away from this analysis is that urban quantities that have some type of social component exhibit increasing returns to scale because cities are social networks of people. The urban quantities subject to the forces of this social network are precisely those urban quantities that depend in some way on social interactions. Thus, the nature of cities, on one hand, allows for higher GMP, income from wages, and patents per capita in larger cities but also results, on average, in higher rates of violent crime and disease per capita. This is an unfortunate fact. I will return to this point after looking at urban quantities that scale sublinearly and urban quantities that scale linearly.

All cities have a social network, but why should each city's social network evolve in such a way that leads cities to scale in such a remarkably consistent way? For the answer to this question, we turn to the generative model introduced by Barabási and Albert (1999) described above. Their model, which assumes only that (1) the network grows and (2) that the addition of new nodes is governed by preferential attachment, can be applied to cities to describe the way they grow. It is tautologically clear that the population growth of cities satisfies (1). Assumption (2), preferential attachment, is also reasonably satisfied by the population growth of cities.

Preferential attachment simply means that new nodes are more likely to be connected to existing nodes with high connectivity than existing nodes with low connectivity. Applied to cities, this is simply the statement that a new resident is likely to seek out people with connections in whatever it is that they may be interested. For example, a young R&D scientist moving to a new city is likely to seek a position that puts her into contact with other people in her field. In other words, she is more likely to establish connections with other R&D scientists (highly connected people in R&D) than she is to establish connections with medical doctors (lowly connected people in R&D). This is preferential attachment, and it seems clear that it offers a reasonable description of the way all cities add new residents. Thus, we can conclude that the Barabási and Albert (1999) model can be applied to the growth of urban social networks. Since Barabási and Albert (1999) proved that their model leads to network scaling, we have arrived at a possible explanation of the urban scaling we have just observed.

We have just seen that the increasing returns to scale exhibited by five superlinear urban quantities can be traced to underlying forces that relate to social interactions, and further that this phenomenon can be described by networks. Now we turn our attention to urban quantities that scale sublinearly, and try to account for the economies of scale that characterize them. Two such urban quantities were analyzed for this paper: the total length of all roadways and the total number of gas stations in a city. A summary of the results is given in Table 2.

Urban Quantity (Y)	β	Std. Error	95% C.I.	Adj-R^2	Obs.	Years
Total roadway length	0.8567	0.006	[0.8450, 0.8685]	0.8349	3434	2001-2008
Gas Stations	0.7063	0.0399	[0.6276, 0.7850]	0.6065	251	2007

Table 2. Scaling exponents for urban quantities that scale sublinearly

The total length of roadways and the number of gas stations in a city are both part of a city's infrastructure. The fact that both scale with exponents less than one means that as cities grow, they require fewer roadway miles and fewer gas stations per capita. In other words, there are economies of scale associated with these urban quantities. Similar findings are reported in Bettencourt et. al (2007). Gregory Ingram and Zhi Liu (1998) also arrive at a similar conclusion, stating that "urban road length has a population elasticity of 0.8, suggesting a decreasing scale effect (p. 12)." A moment's reflection on this result will lead the thinker to consider at least two interesting and fundamental questions: why do these urban quantities behave differently than the superlinear urban quantities discussed above and how can this difference of behavior be explained in terms of networks?

On the first question, the answer, at least to a first approximation, is that there is no clear social component to these two urban quantities. Recall that it was the social nature of the superlinear urban quantities that drove the increasing returns to scale that we observed to be associated with them. Alternatively, there must be some other feature common to sublinear urban quantities that is responsible for the economies of scale that characterize them. If we now consider cities as networks such that roads are the edges and intersections of roads are nodes, simple network theory again offers an elegant explanation of this phenomenon.

This infrastructural network is fundamentally different than the social network associated with superlinear urban quantities described above, and this difference is the source of the divergent behavior of sublinear and superlinear urban quantities. The principal dissimilarity is that the infrastructural network is embedded in physical space, whereas the social network is purely an abstraction. Edges in the infrastructural network of a city are the concrete and asphalt pathways that connect places. Edges in the social network of a city are the invisible relationships and interactions between humans. An obvious implication of this difference between infrastructural and social networks is that there are real, measurable costs associated with building new roads, but no such analogous costs exist in the framework of social networks.

Given the costs associated with adding to a city's infrastructure, it makes intuitive sense for city planners and other officials to seek out the most efficient paths for roadways. However, to simply suggest that larger cities, on average, have less roadway miles per capita because officials are under pressure to create an efficient roadway system is to offer an inadequate, not to mention intellectually dissatisfying, explanation. Furthermore, there is evidence that more fundamental forces are driving this phenomenon. Barthélemy and Flammini (2008) note that "[s]triking statistical regularities [in the urban street patterns of] different cities have been recently empirically found, suggesting that a general and details-independent process may be in action (p. 1)." In other words, there must be some intrinsic property of networks driving the economies of scale that characterize infrastructural urban quantities.

Urban population growth can occur in two ways: new residents can cluster into the space already occupied by a city, or the perimeter of the city can expand to incorporate neighboring areas. The first way leads to an increase in population density, while the second leads to urban sprawl. If the first situation occurs, it is clear that per capita roadway length will fall, because the city will include more people but roughly the same amount of roadways. Ingram and Liu (1998) find empirical support for this observation:

> [M]ost of the increases in urban road length stem from annexation and not from constructing new roads in built up areas. In fact, relatively little new road length is being constructed in the existing urban areas, presumably because the cost of new rights of way is high in both economic and political terms…The strong implication of urban territory annexations is that urban areas provide road capacity by spreading development over space and not by increasing the density of roads in existing built up areas (p. 13).

Thus, the bulk of new roadway length is a result of the expansion of the perimeter of a city, the second situation described above. In terms of networks, this situation can be expressed as the addition of new nodes and edges to non-clustered regions of the network.

Barthélemy and Flammini (2008) introduced an algorithm that gives rise to roadway networks that closely model many real-world road systems. Their model rests on only a few basic assumptions: that there are outlying locations that need to be connected to the network and that the best way to connect these outlying locations to the network is to construct roads such that the reduction of the cumulative distance of the locations from the existing network is maximized. If a road network evolves according to this model, the end result is that "the typical length from [any location] to the existing road network shortens and scales (Barthéelemy and Flammini, 2008, p. 3)."

The lesson here is that as road networks evolve outward, average distances between the locations connected by the network tend to be minimized in a natural way. Stated differently, it is the nature of some networks to become more efficiently connected as they grow. The data suggests that the roadway networks of cities are constructed in a way that this property holds true of them. This is precisely why we observe scaling with regard to the total length of roadways in cities.

Similar reasoning applied to other infrastructural urban quantities can also explain why they scale sublinearly. For example, the other sublinear urban quantity analyzed in this paper is the number of gas stations in a city. Given that there are, on average, fewer roadway miles per person in larger cities, it follows that larger cities will need fewer gas stations on a per capita basis. In larger cities, the roadway system is more efficient, and thus less fuel per person is required to traverse the network of roads. A third infrastructural urban quantity has been found to scale sublinearly by Bettencourt et. al (2007): the total length of all electrical cables in a city. Their results indicate a scaling exponent of $\beta = 0.87$ for this urban quantity. It is reasonable to believe that networks of electrical cables share many of the same properties as networks of roads, because much of a cities electrical infrastructure follows the same layout as its roadway network.

As we have seen, the network of infrastructure of a city is dominated by efficiency and optimization effects as the network grows. This fact means that larger cities are able to enjoy the cost-saving advantages associated with such networks. As with the social network of city, however, there is a downside to this narrative, particularly as it pertains to networks of roadways. Larger networks of roads are more efficient in terms of the distance along the network that it takes to connect locations. Less roadway miles per capita, however, means that there is likely to be more traffic congestion in larger cities. Thus, there are pecuniary cost savings on a per capita basis associated with larger networks of roads, but this is offset to some degree by higher congestion costs. The economic literature is rife with theoretical and empirical work on traffic congestion and its causes and consequences (see, for example, Glaeser, 2011).

The final class of urban quantities consists of those urban quantities that scale linearly. Three such urban quantities were tested for this paper: total employment, property crime, and total housing units. Summary results for linear urban quantities are presented in Table 3. Each of these magnitudes has a scaling exponent roughly equal to one. Thus, these quantities tend to increase one-to-one with city population. Whereas sublinear and superlinear urban quantities were found to exhibit economies of scale and increasing returns to scale, respectively, because of underlying network dynamics, linear urban quantities scale on a one-to-one basis precisely because they are not subject to the properties of networks.

Urban Quantity (Y)	β	Std. Error	95% C.I.	Adj-R^2	Obs.	Years
Total employment	1.0184	0.0021	[1.0148, 1.0226]	0.9819	3294	2001-2008
Property Crime	1.0195	0.0070	[1.0057, 1.0331]	0.9254	1781	2005-2010
Total housing units	0.9811	0.0031	[0.9751, 0.9871]	0.9898	642	2000, 2010

Table 3. Scaling exponents for urban quantities that scale linearly

So far, we have seen how cities can be thought of as social networks of people in the abstract and as infrastructural networks in space. Neither of these networks exerts its influence on linear urban quantities as cities grow. The power of the social network of a city lies in the fact that the upper bound for the number of interactions between people grows faster than the population, and the power of the infrastructural network lies in its tendency to evolve in a naturally efficient way. However, neither of these effects has any bearing on urban quantities that scale linearly.

Total housing units and total employment are urban quantities that reflect the needs of individuals. Individuals make employment and housing decisions based on their needs and resources and the needs and resources of their immediate families. These needs and resources are largely independent of a city's social network. Certainly, these magnitudes are not subject to the properties of infrastructural networks.

The divergent behavior of property crime and violent crime with regard to city size perfectly illustrates this point. Violent crime was found to scale superlinearly because the number of possible interactions increases faster than population increases. Property crime, however, does not respond to this network dynamism. The acts of vandalism and burglary-both of which constitute property crime- do not necessarily entail an interaction between two individuals. Every act of violent crime, on the other hand, requires at least one perpetrator and one victim. Thus, each possible connection between nodes represents a possible perpetrator/victim interaction. There is no such corresponding interpretation in terms of networks for property crime.

In simplest terms, violent crime involves a person and at least one other person. Property crime involves a person and whatever object that person chooses to vandalize or burgle. Violent crime is fundamentally social; property crime is not. Hence, violent crime, because it is subject to the properties of a city's social network, scales superlinearly, but property crime, because it is not subject to those properties, scales linearly.

Large cities are prone to high rates of innovation and infrastructural efficiency precisely because they are networks of people, roads, pipes, and wires. But, as we have seen, the news is not all good. As Ed Glaeser puts it, "the same density that spreads ideas can spread disease (2011)." In other words, the dynamics that are the source of many urban advantages are also the source of many urban ills. Perhaps developing a deeper understanding of cities in terms of networks will allow us to craft policies and institutions that address the ills without eliminating the advantages. But this type of theory of the city is a long way off.

VI. Conclusion

The empirical evidence presented herein combined with previous results- most notably in Bettencourt et al (2007)- provides compelling support for the conjecture that cities scale. We have seen that city size as measured by population is a remarkably powerful predictor of a wide array of urban quantities and that we can classify urban quantities according to their scaling behavior. But to this point, these empirical observations remain just that: observations. Little can be definitively said about the underlying causes of urban scaling.

With that in mind, I have tried to take some small steps toward explaining this phenomenon in terms of the theory of networks. I have suggested that cities are fundamentally composed of two networks, one social and one infrastructural. Furthermore, I have proposed that the fact that these networks evolve according to different processes can explain why urban quantities subject to the properties of each network exhibit divergent scaling behavior. Social networks grow in a way that affords people more opportunities to connect and interact; infrastructural networks grow in a way that tends to naturally increase efficiency. Additionally, I have noted that there are two sides two each of these stories. The same forces that lead larger cities to have higher levels of productivity and infrastructural cost savings also lead to higher rates of disease, violent crime, and traffic congestion.

What I have done is essentially to show that simple network theory aligns with the empirical observations presented here and elsewhere in such a way that networks are a plausible explanation for what we observe in the data. Obviously, much more needs to be done. Specifically, if we are to take the cities-as-networks conjecture to be the true explanation of urban scaling, a full theory is needed. In my view, this theory would at minimum provide answers to the following questions:

- What properties of networks dictate the value of the scaling exponents? For instance, why do the scaling exponents of superliner urban quantities cluster at $\beta \approx 1.1$-1.3 and not some other value?
- In what way do local factors interact with network forces? Stated differently, what effects can we attribute to local policies, laws, and regulations if viewed from within the framework of networks?
- Is there some critical population point at which we can begin to appeal to network theory for answers and below which network forces are not a factor?

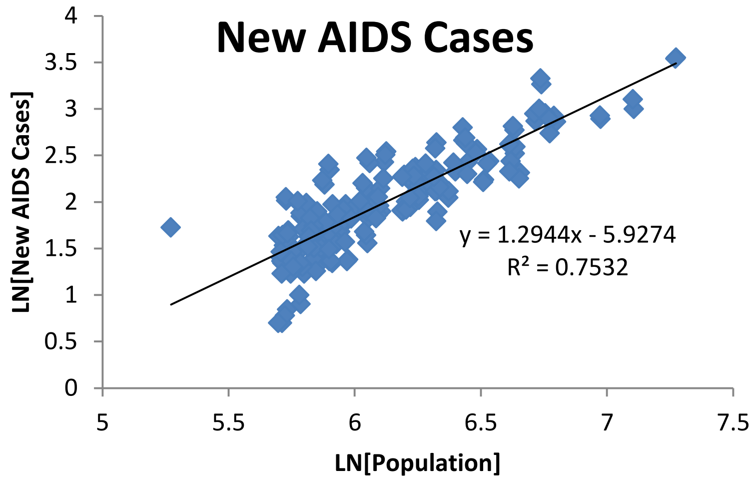
Surely such a theory of cities would be useful. In fact, just observing that cities scale due to forces largely beyond anyone's control can be valuable. One such example of utility involves the development of scale-free metrics that can be used to eliminate size bias when ranking cities. Such metrics, called Scale-Adjusted Metropolitan Indicators (SAMIs), have been proposed by Bettencourt et al (2010). SAMIs rank cities in terms of a particular urban quantity by their residuals with regard to the OLS regression line generated from the data. Bettencourt et al (2010) show that SAMIs tend to minimize scale bias in city rankings and thus provide a clearer depiction of how cities stack up in terms of crime rates or other urban quantities.

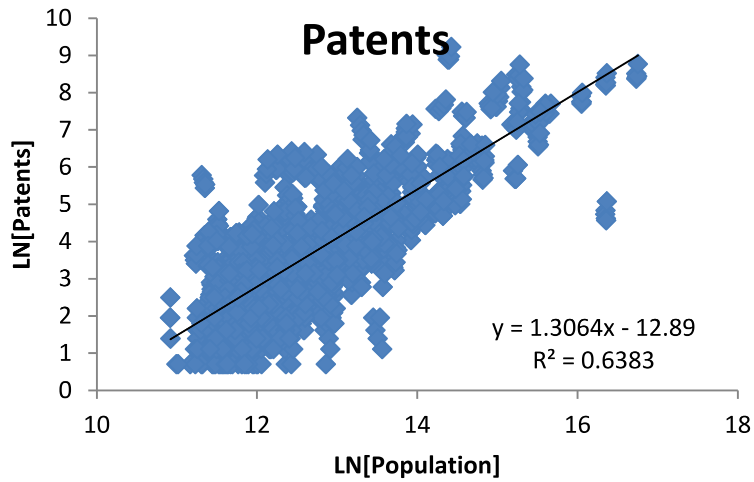
Furthermore, this knowledge can be used to set realistic policy goals. Given that a higher rate of violent crime is a natural consequence of city size, it is not reasonable to expect New York City to achieve a per-capita violent crime rate comparable to that of State College, Pennsylvania. This is not to say that officials should simply throw their hands up and declare that the situation is out of their control. Rather, they could frame the problem in terms of networks and seek to create policies that take into account the social connections that are crucial to the spread of disease and the propagation of violent crime.

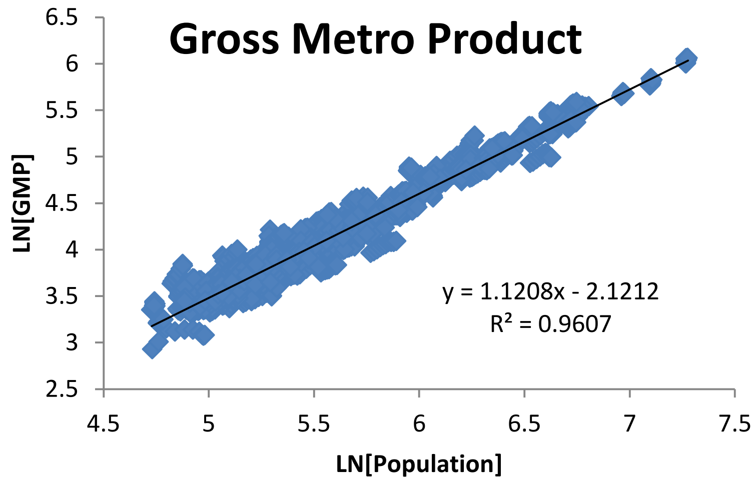
A more rigorous approach to this topic would require that the actual social and infrastructural networks of cities be modeled and then allowed to evolve in simulations according to their respective proposed generative models. The results could then be checked against data to determine whether this idea holds any predictive value. This type of modeling has been done for other types of networks, such as networks of web pages and networks of and references in academic literature (see Barabási and Albert, 1999) and networks of roads (see Barthélemy and Flammini 2008) . However, such an approach for cities is beyond the scope of this paper and, frankly, beyond the current abilities of its author.

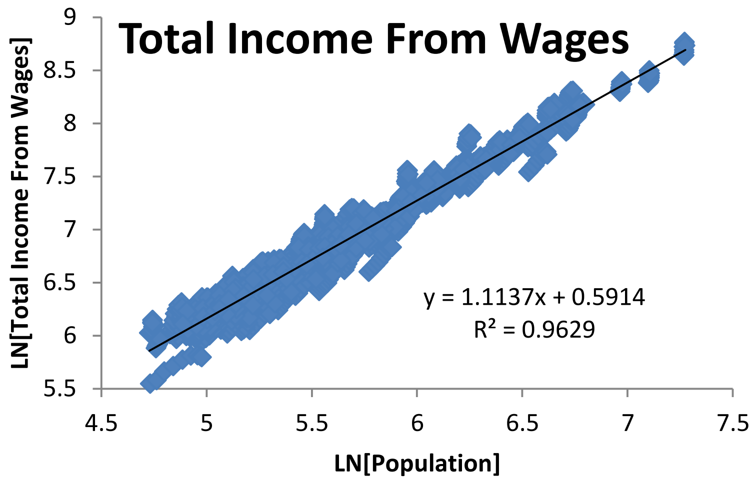
Appendix

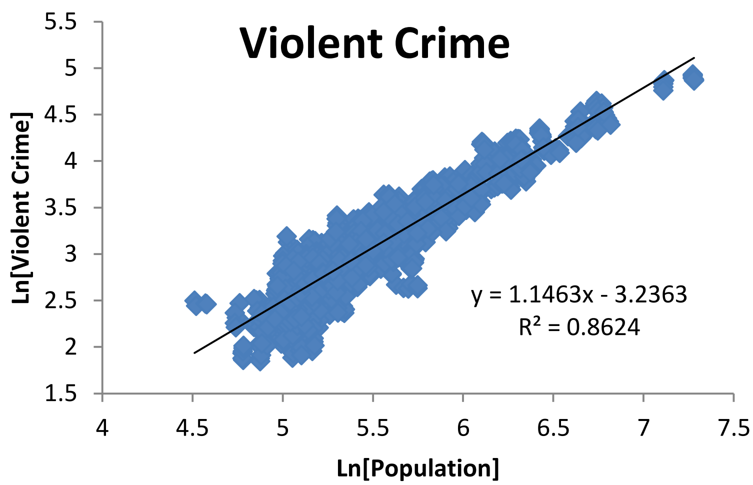
Superlinear Urban Quantities



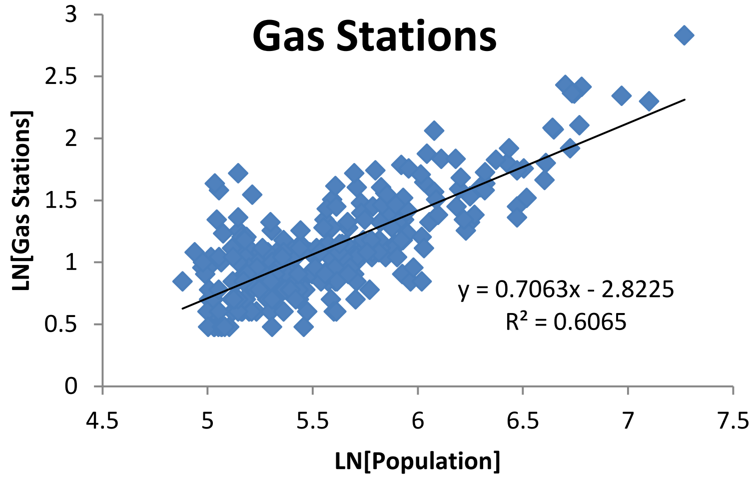


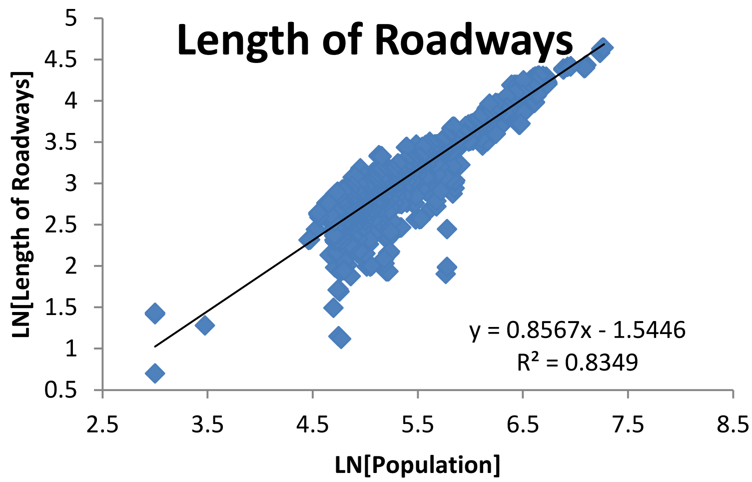



Total Income From Wages
LN[Total Income From Wages]
9
8.5
8
7.5
7
6.5
6
5.5
4.5
5
5.5
6
6.5
7
7.5
LN[Population]
y = 1.1137x + 0.5914
R² = 0.9629

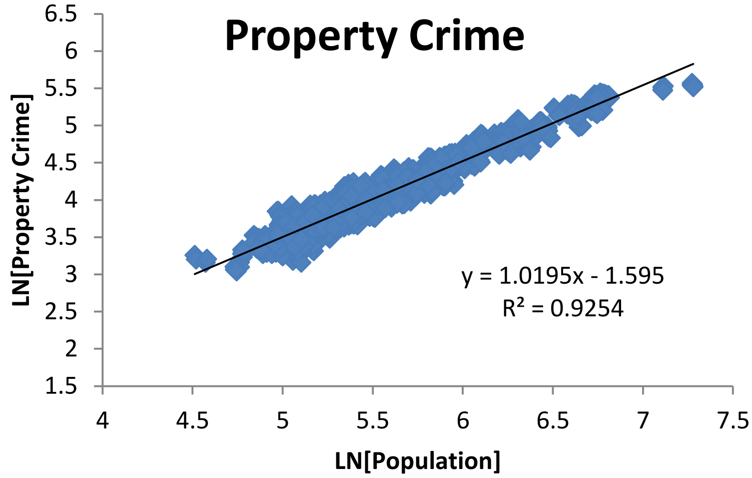

Violent Crime
Ln[Violent Crime]
5.5
5
4.5
4
3.5
3
2.5
2
1.5
4
4.5
5
5.5
6
6.5
7
7.5
Ln[Population]
y = 1.1463x - 3.2363
R² = 0.8624

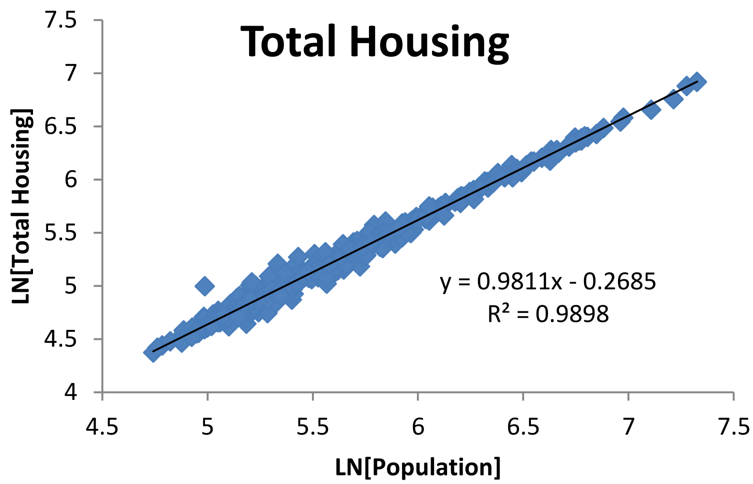
Sublinear Urban Quantities

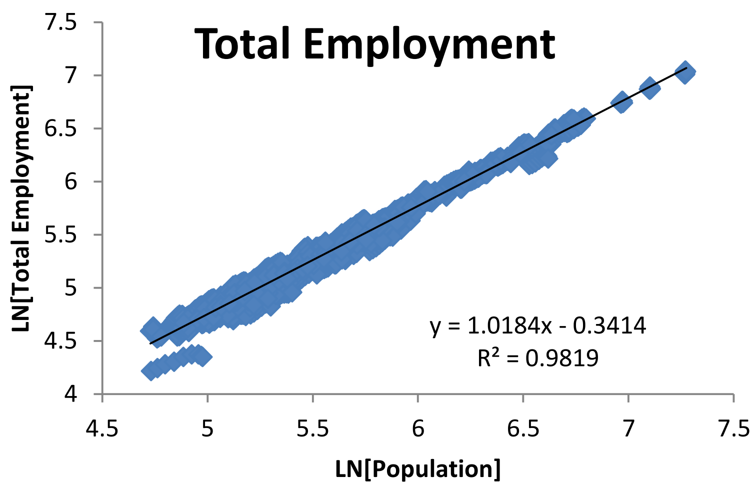


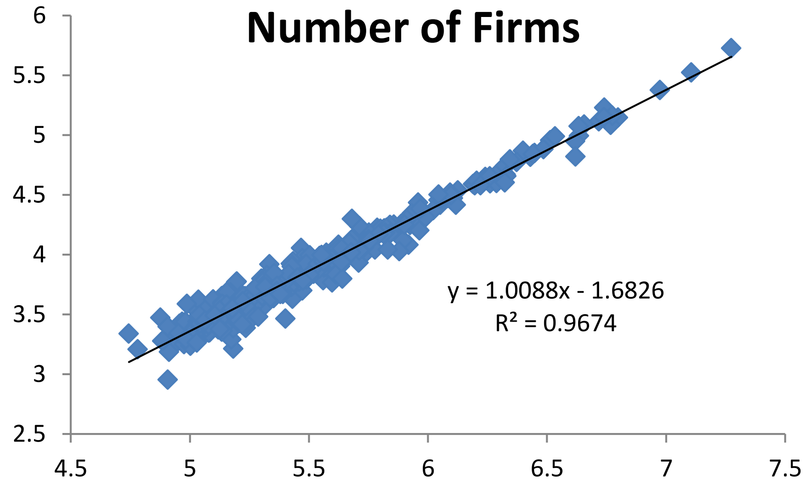


Linear Urban Quantities









Urban Quantity (Y)	β	Std. Error	95% C.I.	Adj-R^2	Obs.	Years
Number of Firms	1.008	0.0088	[0.9914, 1.0262]	0.9674	363	2007

Scaling exponent for total number of firms

References

Albert, R., & Bárabasi, A.L. (1999). Emergence of Scaling in Random Networks, *Science*, 286(5439), 509-520.

Audretsch, D. B., & Feldman, M.P. R&D (1996). Spillovers and the Geography of Innovation and Production. *The American Economic Review*, 86(3), 630-640.

Barthélemy, M. & Flammini, A. (2008). Modeling Urban Street Patterns. *Physical Review Letters*, 100(13), 1-4.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kuhnert, C., & West, G. B. (2007). Growth, Innovation, Scaling, and the Pace of Life in Cities. *Proceedings of the National Academy of Sciences*, 104(17), 7301-7306.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kuhnert, C., & West, G. B. (2010). Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. *PLoS ONE*, 5(11), e13541.

Bettencourt, L. M. A., Lobo, J., & West, G. B. (2008). Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics. *The European Physical Journal B*, 63, 285-293.

Crane, P., & Kinzig, A. (2005). Nature in the Metropolis. *Science*, 308(5726), 1225.

Erdös, P., & Rényi, A. (1960). On the Evolution of Random Graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(17), 17-61.

Federal Highway Administration. *Highway Statistics Series.* Washington, DC. Retrieved from www.fhwa.dot.gov/policyinformation/statistics.cfm.

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics*, 1(1), 255-294.

Glaeser, E. L. (1998). Are Cities Dying? *The Journal of Economic Perspectives,* 12(2), 139-160.

Glaeser, E. L. (2011). *Triumph of the City: How our Greatest Invention makes us Richer, Smarter, Greener, Healthier, and Happier*. London: Macmillan.

Glaeser, E. L., & Sacerdote, B. (1999). Why Is There More Crime in Cities? *Journal of Political Economy,* 107(6), S225-258.

Ingram, G. K., & Liu, Z. (1998). Vehicles, Roads and Road Use: Alternative Empirical Specifications. *Policy Research Working Paper 2036*, World Bank.

Johnson, S. (2010). *Where good ideas come from: The Natural History of Innovation*. New York, Riverhead Books.

Marshall, A. C. (1961). *Principles of Economics*. London: Macmillan.

Rosenthal, S. S., & Strange, W.C. (2006). The Micro-Empirics of Agglomeration Economies. In R.J. Arnott & D.P. McMillen (Eds.), *Blackwell Companion to Urban Economics* (pp. 7-23). Malden, MA: Blackwell.

Smith, A. (1937) *The Wealth of Nations*. New York: Modern Library.

Stagliano, E. (2011, August 24). Millenium Science Complex to Open Soon. *The Daily Collegian*, p. A1.

Stull, W. J. (1986). The Urban Economics of Adam Smith, *Journal of Urban Economics,* 20(3), 291-311.

Sveikauskas, L. (1975). The Productivity of Cities, *The Quarterly Journal of Economics*, 89(3), 393-413.

U.S. Census Bureau. *American FactFinder*. Washington, DC. Retrieved from http://factfinder.census.gov.

U.S. Census Bureau. (2011). *Population Distribution and Change: 2000-2010*. 2010 Census Briefs. Washington, DC. Retrieved from www.census.gov/prod/cen2010/briefs.

U.S. Department of Commerce. *Bureau of Economic Analysis Interactive Data*. Washington, DC. Retrieved from www.bea.gov.

U.S. Department of Commerce. *U.S. Patent and Trademark Office Data*. Washington, DC. Retrieved from www.uspto.gov.

U.S. Department of Housing and Urban Development. *State of the Cities Data Systems*. Washington, DC. Retrieved from www.huduser.org/portal/datasets/socds.html.

West, G.B. (2011). *The Surprising Math of Cities and Corporations*. TED Talks Series. Retrieved from www.ted.com.

Academic Vita

Dane R. Wilburne
47 North Pine St.
Lewistown, PA 17044
drw5010@psu.edu

Education:
B.S. Economics, The Pennsylvania State University Schreyer Honors College, Spring 2012
Honors in Economics

Thesis Title: *Toward a Theory of Cities: Empirical Evidence of and a Proposed Explanation for Urban Scaling*

Thesis Superviser: Dr. N. Edward Coulson
Thesis Adviser: Dr. James R. Tybout

Research Experience:
Research Experiences for Undergraduates Program, Spring 2011

Served as research assistant for Professor Robert C. Marshall, Head Department of Economics

Results of research appear in *The Economics of Collusion: Cartels and Bidding Rings*, by Robert C. Marshal (The Pennsylvania State University) and Leslie M. Marx (Duke University)

Scholarships and Awards
McVey Trustee Scholarship, Fall 2010, Spring 2011, Fall 2011, Spring 2012