Quantifying Creative and Inventive Activities in Urban Areas through Their Full Probability Distributions

Andrés Gómez-Liévano[1,*], Luís M.A. Bettencourt[2], Kevin Stolarick[3], Deborah Strumsky[4], and José Lobo[5]

1 Center for International Development, Harvard University, Cambridge, MA, USA.

2 Santa Fe Institute, Santa Fe, NM, USA.

3 India Institute for Competitiveness, Toronto, Canada.

4 Santa Fe Institute, Santa Fe, NM, USA.

5 School of Sustainability, Arizona State University, Tempe, AZ, USA.

$*$ E-mail: Corresponding agomezlievano@asu.edu

Abstract

How much more creative or inventive can a city be relative to other cities? This question has not been properly framed in previous works due to a lack of knowledge about the probability distribution that describes the levels of inventive and creative activities in cities. It is an important question since differentials in urban creative and inventive activities greatly explain regional differences in economic development. The reason to use probability distributions is because they characterize the full variability of outcomes, and thus, it is the way to quantify the likelihood of observing each level. For US Metropolitan and Micropolitan Statistical Areas, we report here for the first time that lognormal probability distributions describe the counts of inventors and creative employment for urban areas with similar population sizes, and that the unconditioned distribution of inventors and creatives is Pareto with a right-tail exponential cutoff. Quantifying these statistical distributions, moreover, suggests some general mechanisms driving the accumulation of inventive and creative activities in cities, and how these mechanisms change with city population size.

1. INTRODUCTION

It has become a truism that the generation and exchange of knowledge have become the main drivers of economic development through its effects on productivity improvements and technological change (Romer, 1986, 1990; Lucas Jr., 1988; Barro and Sala-i Martin, 2003; Weil, 2012). The principal agents responsible for the generation, recombination and transmission of knowledge, the bearers of "human capital", are, of course, individuals (Lucas Jr., 1988). There is a virtuous cycle between increasing the number of individuals engaged in the generation of ideas and economic growth (Lucas Jr., 1988; Kremer, 1993; Jones, 1995, 2002; Jones and Romer, 2010). The role of individuals as carriers of knowledge is also emphasized by the research highlighting the economic importance of "know-how" and "routines": the tacit knowledge, competencies, skills and experiences which make it possible to integrate technical knowledge and perform specific tasks (Nelson and Winter, 1982; Dosi et al., 1994; Hausmann and Rodrik, 2003; Hausmann and Hidalgo, 2011). Consequently, and in seeking to explain differentiated economic performance across urban areas, much research effort has been devoted to identifying the factors which help determine the concentration of highly-educated, skilled, technically competent, creative and inventive individuals in cities (Piore and Sabel, 1984; Storper, 1997; Acs, 2002; Bettencourt et al., 2007; Andersson et al., 2007; Lobo and Strumsky, 2008; Florida et al., 2008; Bacolod et al., 2009; Gabe and Abel, 2012; Abel et al., 2012; Lobo et al., 2014).

Yet, despite the great analytical efforts exerted trying to understand the agglomeration of human capital—which indeed exhibits great spatial variability—and how we have learned as to what matters and what does not, it is surprising how little we still actually know (Agrawal, 2003; Cooke et al., 2011; Shearmur, 2012). What if the search for a set of prime urban factors (or socioeconomic and cultural features) responsible for the concentration of human capital, invention and innovation is somewhat misguided (Iammarino, 2011)? What if it is unlikely for any one set of location-specific characteristics to uniquely generate an urban environment conducive to innovation? And what if, despite substantial evidence that different factors are responsible for the accumulation of human capital in different cities and at different times, the manner in which these processes interact is the same? In such a case an adequate understanding of urban innovation would have to become more historical and evolutionary, and the expectations for the success of policy-intervention (which usually aim at increasing the supply of a few "key" inputs) should become somewhat tempered.

Building on the discussion in Iammarino (2011), who emphasizes the variety of sources and paths to innovation, we propose a different approach to the study of agglomeration of innovative capabilities in urban areas. We believe that in order to describe the type of processes which generated the observed agglomerations, we need to take seriously the stochastic nature of the metrics typically used as indicators of urban innovative capabilities (or outputs). Identifying the distribution governing a random variable is a first step in this direction, and is the focus of the present paper. Crucially, though, characterizing a statistical distribution accomplishes more than just specifying the range of possible values the variable can assume; it also informs us about the nature of the process that generated the observed data (Limpert et al., 2001; Frank, 2009; Frank and Smith, 2011; Frank, 2014; Sornette, 2012). Suppose we are given the full set of factors that determine the levels of inventors (or any other proxy measure for generators of ideas) across urban areas. But suppose further that we are not, however, given the functional form specifying how these factors come together so as to determine the levels. How could we go about revealing the way they interact? The number of functional forms that can combine all the relevant factors, in different ways, is so large that empirically determining which of the possible functional forms has the best fit is unfeasible. Characterizing the statistical distribution of innovative individuals, however, can tell us (approximately) about they way in which the underlying factors interact. In fact, a statistical distribution has information about how factors interact even if we are not able to fully stipulate the complete set of relevant factors to begin with.

Standard models of human capital accumulation in urban areas (e.g., Rauch, 1993, Glaeser, 1999, Moretti, 2004, Andersson et al., 2007) are largely silent on questions about statistical distributions because they are developed assuming equilibrium in the labor markets. As a consequence, these type of models do not predict any particular probability distribution for how many inventive (or creative or highly skilled) individual we should expect to observe in urban areas. That is, these equilibrium models do not tell us how likely it is to observe cities that deviate from the equilibrium predictions. Another conceptual barrier for the study of the statistical variation is the excessive focus on averages. On the one hand, models for human capital agglomeration assume equilibrium levels are what we observe on average. And on the other hand, the empirical investigations themselves, through regression analysis, often only estimate average associations (Glaeser and Saiz, 2004; Berry and Glaeser, 2005; Florida et al., 2008; Glaeser, 2011; Miguélez and Moreno, 2013; Davis and Dingel, 2014). Supplementing multivariate regressions with explicit identification of the underlying probability distributions is only an additional step, yet key, if we want to have a better understanding of what is possible, and what can be changed, when implementing public policies.

In brief, our discussion contributes to the empirical literature on urban innovation by quantifying the probability distribution governing the accumulation of human capital (here proxied by “creative” occupations and authors of patent applications). We show that the urban variability in innovative individuals, once population size is controlled for, is well described by a lognormal distribution.[1] Our results thus suggest (i) that there is an underlying *multiplicative* stochastic process affecting the levels of urban innovative activity, (ii) that the number of determining factors involved is large, and (iii) that

[1] A random variable which is lognormally distributed is the end product of a large number of factors interacting with each other multiplicatively. That is to say, factors strongly interacting with each other (see Limpert and Stahel, 2011).

there exist constraints on the variance of the logarithmic levels of inventive and creative individuals. We can infer this because we observe the same distribution (i.e., lognormals) across different scales, and for two different measures of human capital (i.e., creatives and inventors).

Although the probabilistic language in which our results are presented is novel, the conclusions from our results are compatible with previous studies in that we show that the differences between cities are structural (i.e., multiplicative) (Muneepeerakul et al., 2013; Strumsky and Thill, 2013) and lead to divergences between them (Berry and Glaeser, 2005; Lobo et al., 2014). We believe our study about these statistical distributions opens new avenues of research that give more attention to the structure of how factors and people interact in cities, as opposed to the identification of specific factors. This, we also believe, will advance our understanding of how innovation in urban areas operates.

2. Data: Measuring Urban Innovativeness

Our spatial units of analysis are 364 Metropolitan Statistical Areas (MSAs) and 574 Micropolitan Statistical Areas, which together comprise the urban system in the United States. (For the rest of the discussion we will use the term “urban area” and “city” interchangeably.) Both MSAs and Micropolitan Areas consist of core county or counties, with a city at its core plus adjacent counties having a high degree of social and economic integration with the core counties as measured through commuting ties. In the case of MSAs the urban core has a population of 50,000 or more while for Micropolitan Areas the core urban population ranges from 10,000 to 50,000 inhabitants. MSAs and Micropolitan Areas are in effect unified labor markets and spatial entities with high levels of socioeconomic integration. Here we use the 2010 definitions for MSAs and Micropolitan Areas. We apply the same definition to construct the urban areas in the years 2008 and 2009 for inventors (more on this below). Detailed information on how Metropolitan and Micropolitan areas are defined can be found in `http://www.census.gov/population/metro/`. For each urban area we use the population size estimates provided by the U.S. Census Bureau (see `https://www.census.gov/popest/data/metro/totals/2009/` and `http://www.census.gov/popest/data/metro/totals/2013/CBSA-EST2013-alldata.html`).

Ever since the work of Mincer (1958) and Becker (1964) the standard measure of human capital has been simply educational attainment (usually the share of a population with a bachelors degree and above). The difficulties of equating human capital with educational attainment, stemming principally from differences in the quality of the education received by individuals and the differences in the economic relevance among types of schooling, have been amply discussed (see, for example, Mulligan and Sala-i Martin, 2000 and Bode and Villar (2014) [Creativity, education, or what? On the measurement of regional human capital]). Furthermore, educational attainment (or years of schooling) does not fully capture an individuals accumulated experience, nor their creativity, innovativeness, and entrepreneurial capabilities. One influential line of research (Florida, 2002) suggests an alternative measure for human capital, based on occupations, specifically a set of knowledge occupations that make up the creative class. A crucial difference between educational attainment and occupation-based measurements of human capital is that educational attainment is a measure of the supply of talent, while creative employment measures the demand for it: in order for an individual to perform a creative occupation, someone needs to be willing to pay for the persons talent. The demand for talent is not a necessary condition for education-based measures of human capital. While some would claim that education is a prerequisite for future creativity, studies like Smith, Carlsson and Danielsson (1984) show that creativity, education, and skill are all distinct but interrelated determinants of individuals productivity. In this sense, creativity is both a complement and substitute to education and skill. Florida (2002) separates education from creativity, defining education as what you have studied for and creativity as what you do in practice. This does not replace education-based measures of human capital in studies of regional and urban economic development, but should be seen as a complementary measure (see, e.g, Florida, Mellander and Stolarick 2008; Lobo, Mellander, Stolarick and Strumsky, 2014).

A different measure of human capital is implicit in the huge research attention bestowed on patenting. One type of intellectual activity with important consequences for technological and economic development is inventionthe creation of new artifacts, methods, processes and materialsand one type of invention, that which results in the granting of a patent, has become a widely used metric for studying the knowledge economy (see, for example, Mansfield, 1986; Griliches, 1990; Jaffe et al., 1993; Acs et al., 2002; Marx et al., 2009). The granting of a patent heralds the arrival of a new process, method, machine, manufacture of composition of matter (the categories of inventions eligible for the protection of a U.S. patent). A salient characteristic of patenting activity in the United States is that it has been an urban phenomenon (Ullman, 1958; Sokoloff, 1988), and remains so today with approximately 93% of all patents granted by the U.S. Patent Office authored by inventors residing in metropolitan areas. Patent analysis has therefore become a well-established framework for investigating locational and spatial aspects of technological advance with much effort having been devoted elucidating the determinants of urban patenting productivity (see, for example, Acs et al., 2002; Bettencourt et al., 2007; Carlino et al., 2007; Lobo and Strumsky, 2008). Patents are generated by inventorsto study locationally-specific invention (proxied by patents) is therefore to study the agglomeration of one type of skilled and creative individuals, namely inventors. And through the information avialable through granted patents, patent authors can be precisly allocated to Metropolitan and Micropolitan Areas.

To construct the counts of creative occupations employment, as defined in (Florida, 2004, Appx. A), we use 2010 employment data from the U.S. Department of Labor's Occupational Employment Statistics (OES) that are available at the metropolitan area level, together with the 5-year estimates of the data from the U.S. Census American Community Survey (ACS 06-10) available at the county level, which we aggregate into micropolitan areas. Data for inventors were obtained from coding inventor's addresses obtained from the U.S. Patent and Trademark Office (USPTO).[2] Despite all the publicly available information about each patent, no unique identifiers are used for inventors. However, using a combination of conditional matching algorithms, it is possible to identify patents' inventors, and locate them geographically. Details about the algorithm used can be found in Marx et al. (2009).

3. Estimations of Probability Functions

Strictly speaking, there is an infinite number of different stochastic processes that can generate any particular probability distribution describing our random variable Y. Thus, observing a particular statistical distribution does not immediately and uniquely determine the underlying generative mechanism. However, one can make reasonable inferences with a proper understanding of the system under study. In the context of cities, the quantities we wish to understand here presumably arise from the aggregation of many processes and many factors. Hence, some limit theorems may apply, and we may observe some of the common limiting distributions.

In our study of urban creative and inventive activities, we will broadly distinguish between *additive* and *multiplicative* stochastic processes. The former refers to the situation when factors are in general acting separately to determine the value of the variable Y, a situation which can be mathematically expressed as $Y = \sum_i F_i$, where F_i is the random effect of factor i. If this is the case, we expect the random variable Y to be normally distributed by invoking the Central Limit Theorem (CLT). In contrast, if factors are acting interactively, which can be represented as $Y = \prod_i F_i$, we expect the outcome Y to be *lognormally* distributed, which can be understood by using the CLT once we apply logarithms to both sides. The standard assumptions behind CLT are that F_i must be independent and identically distributed, each have a finite variance, and the number of terms i must be very large. However, normal and lognormal distributions can still arise in more general situations, given that in these processes of aggregation different forms of information are dissipated, maintained, or amplified (Jaynes, 2003; Frank, 2009; Frank and Smith, 2011). It turns out that whereas additive processes dissipate information about the variance of the individual factors F_i, multiplicative processes amplify them. This, we will show, will be a way to understand why cities diverge from each other in their creative and inventive activities as they grow in population size.

Our analytical attention is focused on the full statistical distributions of the counts of creative employment and inventors. We quantify the distributions in two ways. First, we characterize the distribution of total counts across all urban areas by estimating the marginal distribution $P(Y)$, where Y will refer either to counts of creative employment or to counts of inventors. And second, we characterize the conditional probability distribution of the counts Y *given* population size, that is we estimate we

[2] http://www.uspto.gov/.

estimate $P(Y|X)$, where X is another urban attribute. [Andres: a bit here on the reason/logic behind doing this.] Specifically we condition against urban population size, N, so the conditional relationship becomes $P(Y|N)$. The decision to condition population size recognizes the importance of urban scale as both a consequence and determinant of socioeconomic dynamics (some references here – Jose will provide). This approach has already been pursued in similar contexts recently (Bettencourt et al., 2010; Gomez-Lievano et al., 2012; Alves et al., 2013a,b, 2014; Mantovani et al., 2013), and we apply it here to the study of the sources of creativity and invention.

A set of cities and their associated indicators can be viewed as a statistical ensemble of realizations (y_i, n_i) with $i = 1, \ldots, m$, of the random variables Y and N. We stress again that Y represents counts of either creative workers or inventors, while N represents population size. Since both measures represented by Y count people, they are logically bounded by the population size of the city, i.e., $Y \leq N$. It is a trivial constraint since both creatives and inventors are a subset of the total population, but it underscores, again, the importance of taking into account urban population size explicitly in our analysis. We want to specify the probability of observing $Y = y$ creatives, or inventors, in a city conditional on an urban population of size $N = n$; that is, we want to estimate $P(Y = y|N = n)$, in which the random variables Y and N take values on the non-negative integers. Since we want to understand how probable or improbable a particular level of innovative activity is in an urban area, *given* a certain number of people living and working in it, a necessary component of the analysis is to also quantify the probability across all scales $P(Y = y)$.

For practicality, we will relax the condition that Y and N must take integer values and instead consider them to be continuous. Since our counting of inventors in each year depends upon the existence of patent applications, our numbers are subject to interannual fluctuations. We will use a 3-year average, from 2008 to 2010, to reduce this variation. In the case of creative employment, more than 99 percent of the count numbers are larger than $1,000$ and span a range of more than three orders of magnitude, so the use of a continuous approximation is also valid. This way, the granularity of the data becomes less evident since $P(Y = y) \to 0$. The approximation is even more valid for population size. Hence, we will be estimating probability *density* functions (pdfs) which we will denote by $p_{Y|N}(y \mid n \,;\, \theta_n)$ and $p_Y(y)$, as opposed to probability *mass* functions. We write θ_n to make explicit the fact that the parameters of the conditional pdf are in principle functions of population size n.

All parameters are estimated using Maximum Likelihood Estimation (MLE). To estimate $p_Y(y)$ we propose a parametric functional form (see Equation (6) in the Results section), and we implement the methodology presented in Clauset et al. (2009) to fit it to the data. The shape of $p_Y(n)$ is hard to determine exactly, especially in the tail. We propose a function that is simple, but which provides an accurate characterization of how are the creative and inventive activities distributed across all the different size scales.

To estimate $p_{Y|N}(y \mid n \,;\, \theta_n)$ we first use a logarithmic binning across population sizes. That is, we estimate $P(Y|n_j \leq N < n_{j+1})$, such that $n_{j+1} = an_j = a^{j+1}n_{\min}$, where $n_0 = n_{\min}$, and $a > 1$ determines the bin sizes. We denote the j-th bin as $B_j = [n_j, n_{j+1})$.

The conditional distributions describe the statistical variation of our variables Y for cities of comparable population sizes. To compare the conditional distributions across sizes we define the transformed variables

$$z = \frac{\ln\left(y_{|B_j}\right) - \widehat{\mu_j}}{\widehat{\sigma_j}}, \tag{1}$$

where, for each set of cities with populations within bin B_j, we have calculated the (unbiased) sample mean $\widehat{\mu_j}$ and standard deviation $\widehat{\sigma_j}$ of the logarithm of the corresponding counts of the variable Y in each bin, denoted by $y_{|B_j}$:

$$\widehat{\mu_j} = \frac{1}{|B_j|} \sum_{n_j \in B_j} \ln\left(y_j\right), \tag{2}$$

and

$$\widehat{\sigma_j} = \sqrt{\frac{1}{|B_j| - 1} \sum_{n_j \in B_j} \left(\ln\left(y_j\right) - \widehat{\mu_j}\right)^2}. \tag{3}$$

In Equations (2) and (3), $|B_j|$ denotes the number of observed cities in bin B_j.

The reason to perform this bin-by-bin analysis is that we do not want to presuppose any relationship between Y and N. We are assuming, however, that the function form of the conditional distribution stays constant, which is why we standardize in Eq. (1).

We can visualize the conditional probability density function $p_{Y|N}(y \mid n)$ by plotting histograms of the data, as is customary.

For $p_Y(y)$, however, this visualization method becomes less useful given that the tails of the distributions are heavy and too noisy, i.e., the tails have very few cities with very large populations. In situations like this, distributions are best visualized using logarithmic scales in both axes, and plotting instead the cumulative distribution, or the complementary cumulative (also “countercumulative”), which is more robust to the noise in the tails. Thus, we will plot the marginal probability $p_Y(y)$ through its complementary cumulative function defined as

$$P(Y \geq y) = \int_y^\infty p_Y(y')\mathrm{d}y'. \tag{4}$$

4. Results

We start our analysis by showing the general relationship between the urban indicators y and n in a scatter plot (Figure 1). Both axes are shown in logarithmic scales. Two observations about this plot

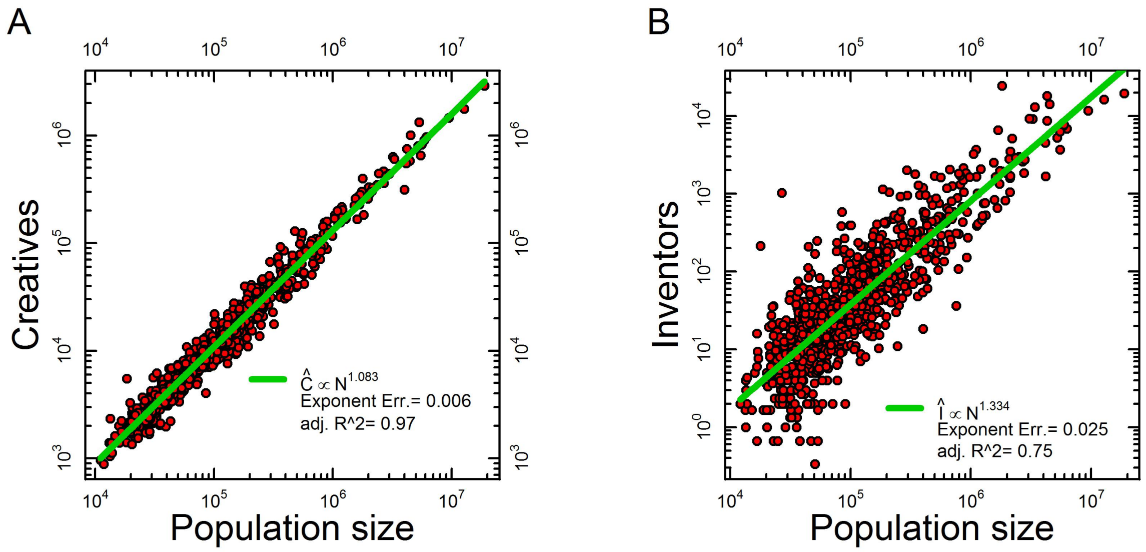

FIGURE 1: Population size scaling of the counts of creative employment C, and inventors I, in US metro and micropolitan areas. In both plots, each point represents a different metro or micro area. The data on creative employment correspond to the year 2010, and for inventors it is the average number over the years 2008, 2009, and 2010. The fitted model is a linear regression on the logarithms of both variables (solid green line). The slope of the regression, thus, is the estimated exponent of the relation $Y \propto N^{\beta}$, where N is population size and Y is C or I. The deviations from this average behavior for the creative employment counts **(A)** show less fluctuations than for inventors **(B)**. These fluctuations are described by the same distribution and suggest the type of process underlying them (see text).

should be noticed: (i) There is broad variability across all the values of y (for both creative employment on the left and inventors on the right), mainly driven by the broad variability in population sizes. This can be best observed for creative employment, for which the relationship has $R^2 \approx 0.97$. And (ii), there is broad variability in y even for cities of the same population size, which is best observed for inventor counts where differences sometimes span more than two orders of magnitude.

In Fig. 1, the green solid line is the linear OLS regression corresponding to the hypothesis that the expected value of Y is given by[3]

$$E(Y|N) = Y_0 N^{\beta}, \tag{5}$$

where Y_0 and β are coefficients whose estimates are shown in Fig. 1.

The observation in Fig. 1 that $\hat{y}_i \propto n_i^{\hat{\beta}}$, where $\hat{\beta}$ is the estimate of the exponent of this power relation (5), is itself interesting. The fact that $\beta > 1$ means that larger urban areas are skill abundant. From a public policy perspective, one may be inclined to propose efforts aimed at increasing creative and inventive activities in cities by increasing their population size. However, as we show in the next section, the estimation of $p_Y(y)$ shows very clearly that the concentration of creative and inventive

[3] A theoretical justification for why average levels of Y are characterized by a power function of N, with exponents clustered around $\beta \approx 1.16$, has recently been given as arising from the spatial mixing of agents, with limited resources and subject to transportation costs, interacting through physical infrastructure (Bettencourt, 2013).

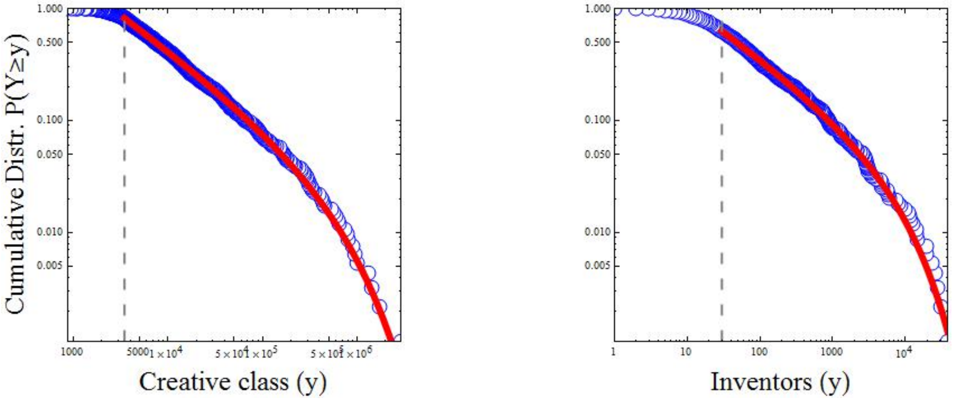

FIGURE 2: Marginal distributions of creatives C and inventors I are well fit by a power-law with an exponential cutoff. We plot the empirical complementary cumulative distribution (blue circles) with the fit (red solid line) corresponding to the function in Eq. (6) using maximum likelihood. The estimated exponents are, respectively from left to right, $\widehat{\tau_C} = 1.623$ and $\widehat{\tau_I} = 1.4603$. The estimates of the characteristic scale of the exponential tail are $\widehat{\gamma_C} = 1\,816\,036.2$ and $\widehat{\gamma_I} = 29\,381.6$. The vertical dashed gray lines are the minimum values for which the distributions hold and are $\widehat{y_{\min C}} = 3491$ and $\widehat{y_{\min I}} = 30$, respectively.

activities becomes in fact increasingly constrained for larger cities.

Figure 2 shows the empirical complementary cumulative distribution of creative employment and inventors in a log-log plot. Qualitatively, the distributions display a scale-free regime modulated by a sharper decay in probability at large population sizes. A simple characterization of this distribution is a power-law with an exponential cutoff:

$$p_Y(y;\tau,\gamma,y_{\min}) = C\frac{\mathrm{e}^{-y/\gamma}}{y^{\tau}}, \quad y \geq y_{\min}, \tag{6}$$

where $C \equiv \frac{\gamma^{\tau-1}}{\Gamma(1-\tau, y_{\min}/\gamma)}$ is a constant of normalization, $\tau > 0$ is the exponent of the power-law, γ is the characteristic scale above which the exponential decay becomes strong, and $\Gamma(z,a)$ is the upper incomplete gamma function[4]

$$\Gamma(z,a) = \int_a^{\infty} t^{z-1}\mathrm{e}^{-t}\mathrm{d}t. \tag{7}$$

Since our data is left-censored we do not fit the lower tail of the distribution. Instead, we consider the values $y \geq y_{\min}$ above a minimum value for which this model holds.

The rationale behind Equation (6) is phenomenological and comes from two sources. First, power-laws (also referred as Pareto or Zipf distributions[5]) in urban sizes have been often assumed as the rule, being the result of processes by which bigger cities attract more people giving rise to "rich-gets-richer"

[4] The lower and upper incomplete gamma functions, $\gamma(z,a)$ and $\Gamma(z,a)$ respectively, are such that $\gamma(z,a) + \Gamma(z,a) = \Gamma(z)$.

[5] A Pareto distribution is a power-law with a density $p(x) \propto x^{-\alpha}$, where the exponent α is larger or equal than two. The case when $\tau = 2$ is special because all its moments are infinite, and is called Zipf's law.

effects that usually lead to such heavy-tailed distributions (Zipf, 1949; Simon, 1955, 1968; Gabaix, 1999; Soo, 2005). Economic foundations of the power-law distribution of population sizes usually rely in Gibrat's Law, whereby the growth of city populations is independent of their size (the original explanation comes from Gabaix, 1999). And second, the exponential part of the distribution comes from the fact that real systems, such as cities, are finite and therefore size and growth have ultimate limits. In physics, such effects are called "finite-size" effects, and are often characterized by such exponential decays (Newman, 2005). These finite-size effects constrain the systematic accumulation of creative and inventive activities for the largest cities, at least for the US system of urban areas.

Based on our proposed functional form, we distinguish distinct regimes in this distribution from comparison to a straight line (in the log-log plot). A straight line in this type of plot would be the signature of a Pareto distribution, and it is indicative of a lack of characteristic scales[6].

According to Figure 2, however, only the middle range of US city sizes displays this scale-free phenomenon. Cities such as A, B, and C belong to this collection of cities. But deviations from the Pareto law in the right tail suggest that scale becomes relevant for large populations. For creatives, this scale is estimated to be $\widehat{\gamma_C} = 1\,816\,036.2$, and for inventors $\widehat{\gamma_I} = 29\,381.6$.

The specification given in Equation (6), and its superior fit over other specifications (see Appendix for details), suggest that there is in fact a scale above which the accumulation of creative employment and inventors, driven by population growth, no longer follows the proportionate growth (Gibrat's Law) implied by the Pareto behavior of the medium sized cities. This conclusion is in contrast with the recent interpretation reported by Berry and Okulicz-Kozaryn (2012), in which the authors claim that a cutoff would be due to a geographic underspecification of these regions.

Figure 1 shows that the average of creative employment and inventors counts varies regularly as a power of population size, characterized by Equation (5). The statistics of the deviations around this relationship will tell us how far, in probabilistic terms, can a city increase, or decrease, its creative and inventive activities. Figure 1 also shows that the behavior of creatives is different from that of inventors, as the latter displays much more dispersion around the regression line. In this section we quantify this behavior by estimating the probability density of these fluctuations.

Conditional distributions. The quantity z (see Eq. (1)) for creatives and inventors is well fitted by a standard normal distribution (Figure 3), which means that the untransformed numbers y_i are lognormally distributed, conditioned on population size.[7] The density of the random variable Y can thus be written as:

$$(8) \qquad p_{Y|N}(y \mid n\,;\, \mu_n, \sigma_n) = \frac{1}{y\sqrt{2\pi\sigma_n^2}} \mathrm{e}^{-\frac{1}{2\sigma_n^2}(\ln(y)-\mu_n)^2}.$$

[6] Pure power-law functions $f(x) = Ax^a$ lack characteristic scales, and are often referred to as "scale-free" functions, since the ratio $f(\lambda x)/f(x) = \lambda^a$ is independent of the scale x.

[7] In this analysis, the instances when $y = 0$ have been excluded.

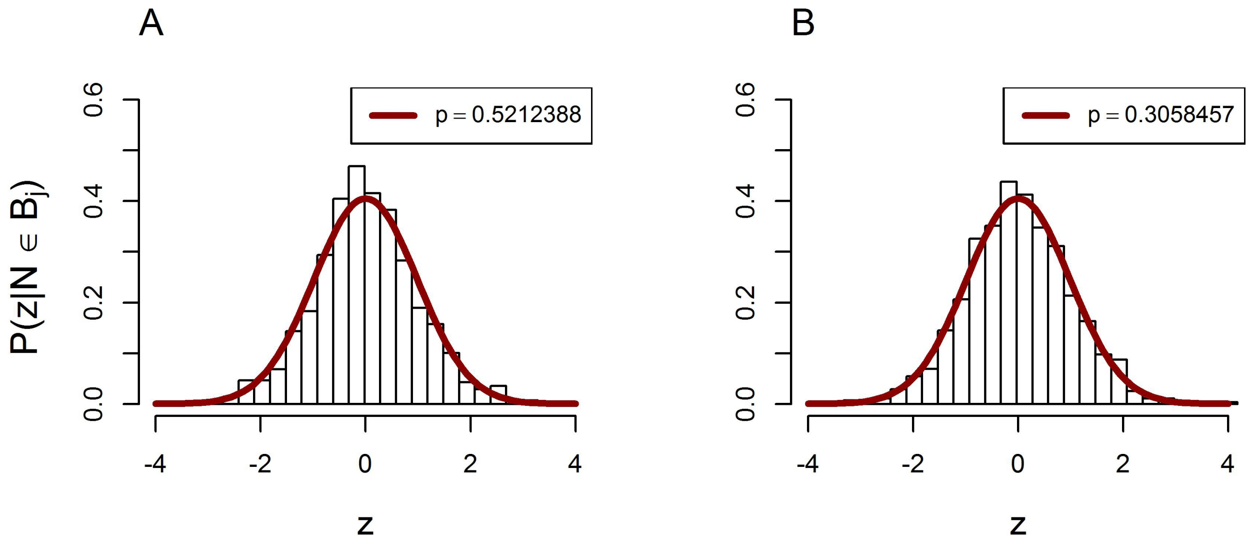

FIGURE 3: Histograms of creative individual counts conditional on population size. We fit a standard normal distribution to the normalized frequency histograms of the standardized logarithmic counts of creatives **(A)** and inventors **(B)**. For both plots the counts were transformed $y_{|B_j} \to z = \frac{\ln(y_{|B_j}) - \widehat{\mu_j}}{\widehat{\sigma_j}}$, where $y_{|B_j}$ stand for the values of Y for cities with $N \in B_j$, and $\widehat{\mu_j}$ and $\widehat{\sigma_j}$ are the corresponding sample mean and standard deviation of the log-counts of that bin. The bin size used to construct the B_j intervals, was set to $a = 1.225$ so that there was a good balance between the number of observations per bin ($\sim$ 25) and the total number of bins ($\sim$ 37). The p-values shown are from a Chi-square goodness-of-fit test, where we consider $p > 0.10$ to be an acceptable level to not reject a normal distribution. See the Appendix for more details on goodness-of-fit tests.

Based on the goodness-of-fit tests presented in the Appendix, we conclude that Equation (8) provides a good description of the number of creative class workers and inventors for cities of comparable population size. .

Lognormal distributions have a long history in economics (Aitchison and Brown, 1957), and in the natural sciences in general (Redner, 1990; Limpert et al., 2001). In this sense, it is not a surprise to find the distribution of creatives and inventors to be lognormal. However, it is an unexpected result given that the variables Y of total counts can be written as the summation over all N individuals in the city,

$$Y = \sum_{k=1}^{N} X_k, \tag{9}$$

where $X_k = 1$ if individual k belongs to the creative class workforce (or to the group of inventors), and $X_k = 0$ if not. Expressing Y as a sum of random variables as in Equation (9) reminds us of the Central Limit Theorem. Since N is a large number, and X_k are random variables with finite variance, the random variable Y would be normally distributed *if the X_k were independent random variables.* Since we do not observe normality in the distribution Y, we conclude that X_k are not independent. In other words, being a creative or inventive individual literally *depends* on whether others in the city are

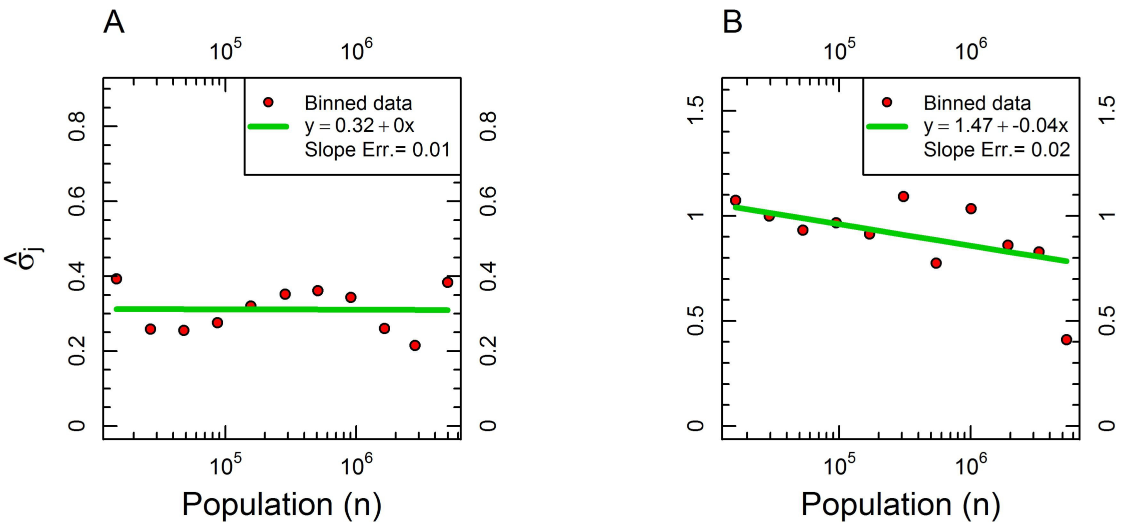

FIGURE 4: Population size dependence of the log-normal parameters estimates of each bin. Creatives **(A)** and inventors **(B)** are shown here to differ in how the variance of the logarithmic counts $\widehat{\sigma_n^2} = \sum(\ln Y - \widehat{\mu_n})^2/(m-1)$ depends on size, and the level. Although the variance is much higher for inventors, there is a weak tendency to decrease with population size.

creative or inventive. Formally, this is

$$p_Y(y) \neq Normal \implies P(X_k = 1, X_{k'} = 1) \neq P(X_k = 1)P(X_{k'} = 1), \tag{10}$$

for two individuals $k, k' = 1, \dots, N$ in the city.

Population size dependence of the log-variance. As was shown in the previous section, each bin corresponds to a collection of lognormally distributed variables. Figure 4 plots the n-dependence of the parameter σ_n of those bins. For creatives $\widehat{\sigma_n} \approx 0.32$ while for inventors $\widehat{\sigma_n} \approx 1.47 - 0.04\ln(n)$, quantifying the fact that creative employment counts vary less than inventor counts, given population size.

The results coming from the behavior of the σ parameters stand as an important restriction on the type of models that should generate our statistics. In particular, they say that the variance of the logarithmic variables varies weakly with population.

The standard deviation $s(Y|N)$ of the lognormally distributed random variable Y is, by definition, proportional to the mean. Specifically,

$$s(Y|N) = \left(\mathrm{e}^{\sigma_N^2} - 1\right)^{1/2} E(Y|N).$$

Using what we know about $E(Y|N)$, we conclude that the standard deviation s of the variable Y scales with population size as

$$s(Y|N) \propto N^{\beta_\sigma}, \tag{11}$$

where $\beta_\sigma \approx \beta$ for creatives (Figure 4A), and $\beta_\sigma \lesssim \beta$ for inventors (Figure 4B). Since the levels of creative and inventive activities are strongly correlated from one year to the next (Bettencourt et al., 2010), increases in population size will exacerbate any existing deviation from the mean.

The parameter μ_n is not shown, but it can be deduced from the fitted line shown in Figure 1. The full specification of $p_{Y|N}(y|n)$ allows us to calculate, for example, what is the most probable level of creative and inventive activities for each metro and micro area according to their population size, and how probable it is that they have their current level, or one greater. Table 1 presents 12 outlier cities (5 micropolitan and 7 metropolitan areas) which had log-deviations from the log-mean, in either creatives or inventors, whose likelihood was less than one over the total number of cities. In mathematical terms, we chose those cities for which $y \notin [\mathrm{e}^{\mu_n - z\sigma_n}, \mathrm{e}^{\mu_n + z\sigma_n}]$. The value of z has been chosen such that $1 - \Phi(z) < 1/938 \approx .00107$, where $\Phi(\cdot)$ is the cumulative normal standard distribution. This yields $z = 3.07$,

Since lognormal distributions are skewed and our sample of cities in each population bin is finite and small, the most probable value $\mathrm{e}^{\mu_n - \sigma_n^2}$ stands as a good comparison point to assess the creative and innovative profile of urban areas.

The Los Alamos micro area is the smallest of the cities, yet has more than 200 times the number of inventors that was most probable to have, and three times as many creatives, according to its size. On the other extreme, the San Jose metro area is the largest of in the Table, and has also very unlikely counts of both inventors and creatives. Likewise, for Durham-Chapel Hill. Table 1 gives quantitative calculations that show how these urban areas are very unlikely cities, and that the reasons behind their extreme values must be analyzed separately.

The cases of Mountain Home (ID) and Clovis (NM) are interesting. They belong in the table because the former has more than 500 times more inventors than it should, and the latter has five times less, which according to their size makes them outliers. But their respective number of creatives is also interesting. The skewness of the lognormal distribution enables the fact that, while having more creatives than the most probable value, they have more than 50 percent likelihood for increasing their creative employment. Palm Coast (FL) presents a similar case: it has more than the probable number with respect to inventors, but less with respect to creatives. In fact, it belongs to the table because it has an extremely low level of creative employment. These three cities, given the whole US urban system in which they are embedded, have good potential to increase their creative activities.

Our main finding is that the statistics of creatives and inventors for cities of comparable population size are all well described by lognormal distributions. One of the implications behind the lognormality of Y is that the distribution is determined by two parameters only. Using tools from information theory, a lognormal distribution can be understood to arise from the maximization of entropy subject to informational constraints on the geometric average and the geometric standard deviation, i.e., e^{μ}

TABLE 1: US urban areas, sorted by population size, that have counts, in inventors or creatives, that are outside a $z = 3.07$ *sigma* interval around the log-mean, i.e., $y \notin [\mathrm{e}^{\mu_n - z\sigma_n}, \mathrm{e}^{\mu_n + z\sigma_n}]$. The numbers shown in the Population and Inventors columns are estimated averages from 2008-2010, although all values representing counts have been rounded to the nearest integer. The value $z = 3.07$ corresponds to a log-deviation such that $1 - \Phi(z) < 1/938$. The random variables and their values are for inventors I and i, for creatives C and c, respectively

Name of Urban Area	**Population (n)**	**Inventors (i)**	**Most probable I**	$P(I \geq i \mid n)$	**Creatives (c)**	**Most probable C**	$P(C \geq c \mid n)$
Los Alamos, NM (Micro)	17,899	213	1	0.	5,502	1,697	0.
Mountain Home, ID (Micro)	26,926	1,027	2	0.	2,801	2,751	0.562
Clewiston, FL (Micro)	39,109	8	4	0.626	1,976	3,848	0.999
Clovis, NM (Micro)	47,009	1	5	0.999	4,504	4,461	0.577
Eagle Pass, TX (Micro)	53,392	1	7	1.	4,328	5,144	0.834
Palm Coast, FL (Metro)	94,755	44	15	0.426	4,040	8,958	1.
Lake Havasu City-Kingman, AZ (Metro)	200,447	43	43	0.82	10,940	21,603	0.999
Merced, CA (Metro)	253,198	37	59	0.924	13,650	27,633	0.999
Ocala, FL (Metro)	330,780	69	87	0.873	17,620	37,218	1.
Durham-Chapel Hill, NC (Metro)	498,511	1,692	155	0.028	128,900	57,900	0.001
McAllen-Edinburg-Mission, TX (Metro)	758,064	36	279	0.999	60,400	87,582	0.963
San Jose-Sunnyvale-Santa Clara, CA (Metro)	1,818,864	24,531	952	0.	396,820	240,275	0.033

and e^{σ}, respectively (Frank and Smith, 2011). This means that the processes that increase or decrease creative and inventive activities in cities are sensitive, in a multiplicative way, to the variations of a large number of input factors.

Lognormal distributions also indicate that the counts of people involved in creative activities increase from a conjunction (product) of effects ("A *and* B *and* ..."), as opposed to normal distributions which arise from disjunction (sum) of events ("A *or* B *or* ..."). The way in which the expressions $Y = \sum_k X_k$ and $Y = \prod_i F_i$ can be mutually compatible is if the factors F_i influence all individuals $k = 1, \ldots, N$ in a similarly multiplicative way. The fact that factors influence all individuals is part of the reason why the random variables X_k are correlated. Given our results, we are unable to say whether factors F_i are endogenous to the social networks of the city, or exogenous as in the case of the weather and the geography.

Regardless, the conclusion is the same. The right regional amenities, job opportunities, partnerships, etc., all have to act in consonance in order to increase the creative endowments of a city. From a public policy perspective, one of the implications is that there is no single-subject silver bullet for fostering innovation in cities of a given population size. Increasing the creative endowments of a city (e.g., creative employment and inventors) requires a coordinated array of propitious circumstances such that all the steps necessary in their enhancement are successful. This is characteristic of multiplicative processes. However, such processes in cities must be further restricted by the population-size (in)dependence of the distribution parameters as discussed above.

5. Discussion

We have characterized the full statistical distribution of creative and inventive activities, measured as counts of individuals, within and across different population size scales, for the U.S. system of metropolitan and micropolitan areas. Three main facts were established in our analysis. First, the counts of creatives and inventors is a Pareto distribution for medium-sized cities, but exponential for the large-sized. Second, for cities of the same population size, the counts are lognormally distributed. And third, both the mean and the variance of the lognormal distribution are power-law functions of population size.

Our focus on probability distributions generates an epistemological question. Are the distributions simply a statement about our ignorance regarding the factors and mechanisms behind the processes of creativity and invention? Or do they reflect a deeper characteristic of how cities work, namely, that such innovation processes in cities are inherently stochastic?

Regarding innovation and the growth of cities, Agrawal notes that "[through] a comprehensive survey of the modern literature on innovation and regional growth, [one] discovers how much we don't know about the mechanisms at work behind the curtains—it is predominantly the statistical correlations between presumed input factors and outputs that assume the spotlight in recent empirical work on this topic" (Agrawal, 2003, p. 460). The assumption behind traditional efforts to understand the sources of innovation is that some few, specific, and well-defined factors underlie the agglomeration of creative and inventive individuals. The presence of lognormals, we have shown, suggest otherwise. By invoking the Central Limit Theorem, lognormals suggest that the number of factors is large. A limitation of our study, at first sight, is that we have only analyzed how likely or unlikely is to observe a city's creative and inventive endowments, eschewing any mention to specific mechanisms. But in fact we believe framing this question in the language of probabilities contributes to our understanding of the "mechanisms behind the curtains".

The empirical distributions reported here suggest two characteristics about the mechanism that determines the levels of creative and inventive activity in cities. One, is that this process is some type of constrained multiplicative random process. And second, that the behavior of individuals is correlated in cities. This evidence of a multiplicative process that correlates individuals strongly suggests, in turn, that the factors influencing creativity and inventivity, positively or negatively, get amplified, and this emphasizes the need to have structural views of the functioning of cities.

In analyzing the processes of innovation, it is possible that we will only understand them if we explicitly deal with the statistical distributions involved, as opposed to only the averages. And reasons abound, in the sense that cities, by virtue of being collections of a diversity of people, firms and institutions, connected through a myriad of physical and informational networks, fluctuate from their expected behavior, often very wildly. Further studies that model cities in a stochastic way are therefore needed.

As the nations draw more heavily on their capacity to produce knowledge to foster economic growth, and as its total population grows and gets increasingly concentrated in cities, questions about how the power of creativity and inventivity can be harnessed to keep creating wealth also become central. Here we have delineated, in precise quantitative terms, the probabilistic landscape in which cities embedded in a larger system can be found.

A clear statistical description of cities across an urban system had been missing as a way to understand innovation in urban systems. Here we have provided some empirical facts regarding this. Still, these facts remain to be explained and reproduced by a detailed mechanistic model. Given the suggestive evidence of multiplicative processes at play, further studies should aim to identify and quantify explicitly the different steps involved in attracting creativity and the production of knowledge.

Our cross-sectional study of the US urban system is still missing an analysis across time. That is, future studies should aim to characterize the distribution of creative and inventive activities for individual cities throughout the years. Bettencourt et al. (2010) highlight the fact that urban aggregate measures, in general, are highly correlated in time. This is relevant to our study since it indicates that

TABLE 2: Comparing probability density function fits for the distribution of creative employment conditioned on population size. This table shows the goodness-of-fit of different bell-shaped standard distributions to the logarithmic counts of the data. The best model, in relation to the others presented, is presented in bold. Note, that there is no parameter estimation involved here, since the logarithmic counts have been standardized

Goodness-of-fit	Laplace(0,1)	Logistic(0,1)	Cauchy(0,1)	Lognormal(0,1)
Log-likelihood	-1354.99	-1492.1	-1531.89	**-1306.53**
AIC	2713.99	2988.21	3067.79	**2617.06**
BIC	2723.66	2997.88	3077.46	**2626.74**
p-value (Anderson-Darling)	$< .0001$	0.	0.	**0.1466**
p-value (Pearson χ^2)	$< .0001$	$< .0001$	$< .0001$	**0.6231**

measures of cross-sectional variation are the result of cumulative effects over time.

ACKNOWLEDGMENTS

We thank Michael Batty, Simon DeDeo, Charlotta Mellander, Oscar Patterson, Geoffrey West, and HyeJin Youn, for useful discussions and comments on parts of this work. AGL thanks the Mathematical, Computational & Modeling Sciences Center (MCMSC) for the financial support and the Martin Prosperity Institute and the Santa Fe Institute for hosting much of the work. AGL also wants to thank Cathy Yang Liu and participants of the 2013 North American Regional Science Council conference for helpful comments on an earlier version of this work. This research was partially supported by the Rockefeller Foundation, the James S. McDonnell Foundation (grant no. 220020195), the John Templeton Foundation (grant no. 15705), the Army Research Office (grant no. W911NF1210097), the Bill and Melinda Gates Foundation (grant OPP1076282) and by a gift from the Bryan J. and June B. Zwan Foundation.

Appendix

Conditional distributions goodness-of-fit tests

In Tables 2 and 3 are shown some comparative tests of different distributions that could be fitted to the conditional distributions of creative employment and inventor counts. Since, from Fig. 3 it is clear that the logarithmic variables have histograms that are bell-shaped, we consider in the analysis four standardized distributions: laplace, logistic, cauchy, and log-normal. The distributions that are not rejected with a confidence level of $p = 0.05$ are shown in bold.

TABLE 3: Comparing probability density function fits for the distribution of inventors conditioned on population size. This table shows the goodness-of-fit of different bell-shaped standard distributions to the logarithmic counts of the data taking. The best model, in relation to the others presented, is presented in bold. Note, that there is no parameter estimation involved here, since the logarithmic counts have been standardized

Goodness-of-fit	Laplace(0,1)	Logistic(0,1)	Cauchy(0,1)	Lognormal(0,1)
Log-likelihood	-1349.19	-1487.82	-1526.17	**-1304.19**
AIC	2702.82	2979.63	3056.35	**2612.39**
BIC	2712.49	2989.30	3066.02	**2622.06**
p-value (Anderson-Darling)	$< .0001$	0.	0.	**0.1357**
p-value (Pearson χ^2)	$< .0001$	$< .0001$	$< .0001$	**0.23059**

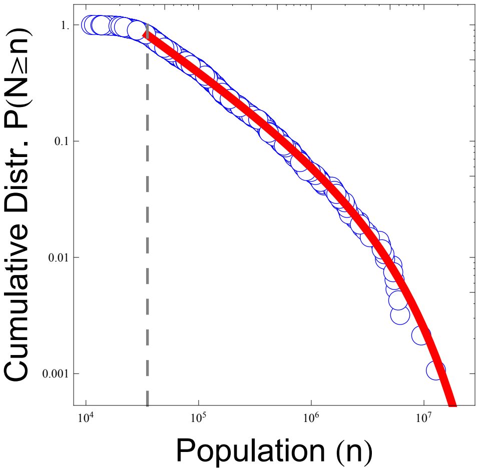

FIGURE 5: Complementary cumulative distribution of population sizes. In this plot it can be seen that a cutoff in the largest population sizes is present, suggesting an accumulated effect from constraints to growth. For fitting the density function (12) of $p_N(n)$, we used the estimated parameters of the density $p_Y(y)$ for creative employment counts (see Appendix). Thus, $\widehat{\beta} = 1.083$, the estimated exponent is $\widehat{\alpha} = 1.675$, the estimated characteristic scale of the exponential tail is $\widehat{\nu} = 11,329,658$, and the model holds for populations above $\widehat{n_{\min}} = 35,141$ (dashed gray vertical line). See Appendix for goodness-of-fit tests.

Marginal distribution fits

Figure 5 shows the empirical complementary cumulative distribution of population sizes. The function

$$p_N(n;\alpha,\beta,\nu,n_{\min}) = C\frac{\mathrm{e}^{-\left(\frac{n}{\nu}\right)^{\beta}}}{\left(\frac{n}{\nu}\right)^{\alpha}}, \quad n \geq n_{\min}, \tag{12}$$

provides a good fit (solid red line). Here, C is a constant of normalization, $\alpha > 0$ stands as an exponent that determines the broadness or narrowness of the distribution, ν is a characteristic scale above which an exponential decay dominates, and β is the scaling exponent in Equation (5). This function supports the widely accepted result that population sizes are well described by a Pareto distribution in the upper tail, but here we find evidence of a cutoff for the largest cities (see Berry and Okulicz-Kozaryn, 2012, Section 4 and the references cited therein for a discussion about this cutoff).

The rationale behind Equation (12) comes from the fact that we actually estimate the marginal distribution of Y, and we use the close relationship between creatives and population size to derive the distribution of N. The close relationship between Y and N (Figure 1) suggests that whichever behavior we see in the distribution of one we will also see in the distribution of the other. It turns out, the distribution of Y shows the usual heavy tailed Pareto behavior seen for population sizes, but with a strong signal of a decay for large numbers. The reason this decay is more easily detected in Y than in N, we argue, is because the exponent $\beta > 1$ strengthens such decay. In the following, we will first estimate the empirical distribution of Y, we will then show the derivation for the distribution of N, and we will present goodness-of-fit comparison with other distributions.

Now, from Figure 1B we know that more than 97 percent of the variability in creative employment is explained by population size. If we assume the relationship $Y = Y_0 N^\beta$ is exact, we can use the conservation of probabilities

$$p_Y(y)\mathrm{d}y = p_N(n)\mathrm{d}n \tag{13}$$

to derive the distribution of N. Since $\mathrm{d}y/\mathrm{d}n = \beta Y_0 n^{\beta-1}$, we have that

$$\begin{aligned} p_N(n) &= p_Y(y)\frac{\mathrm{d}y}{\mathrm{d}n} \\ &= \frac{\beta\gamma^{\tau-1}}{\Gamma(1-\tau, y_{\min}/\gamma)}\frac{\mathrm{e}^{-Y_0 n^\beta/\gamma}}{Y_0^{\tau-1} n^{\beta\tau-\beta+1}}. \end{aligned} \tag{14}$$

Equation (14) can be written as

$$p_N(n;\alpha,\beta,\nu,n_{\min}) = C\frac{\mathrm{e}^{-\left(\frac{n}{\nu}\right)^\beta}}{\left(\frac{n}{\nu}\right)^\alpha}, \quad n \geq n_{\min}, \tag{15}$$

where the constant of normalization is given by

$$C = \beta\frac{\nu^{-1}}{\Gamma\left(\frac{1-\alpha}{\beta}, \left(\frac{n_{\min}}{\nu}\right)^\beta\right)},$$

and

$$\alpha = \beta(\tau - 1) + 1 \tag{16}$$

$$\nu = \left(\frac{\gamma}{Y_0}\right)^{1/\beta} \tag{17}$$

$$n_{\min} = \left(\frac{y_{\min}}{Y_0}\right)^{1/\beta}. \tag{18}$$

This way, although the distribution of N has four parameters, they are fully determined by the parameters of $p_Y(y)$ and the regression between Y and N. When using the estimates $\widehat{\tau}$, $\widehat{\gamma}$, and $\widehat{y_{\min}}$,

TABLE 4: Different estimated probability density functions to fit population size distribution. PLEC stands for "Power-law with exponential cutoff", given by Equation (6), and PL stand for "Power-law", given by a density $p(n) \propto n^{-\alpha}$. All four distribution where truncated from below by the same $\widehat{n_{\min}}$ estimated using Equation (18)

Goodness-of-fit	$p_N(\alpha, \beta, \nu, n_{\min})$	PLEC$(\tau, \gamma, n_{\min})$	Lognormal(μ, σ)	PL$(n_{\min}, \alpha)$
Loglikelihood	**-9996.41**	-10546.3	-10223.5	-10000.6
AIC	**20000.8**	21098.5	20451.1	20005.2
BIC	20019.4	21112.6	20460.3	**20014.5**
p-value (Anderson-Darling)	**0.6889**	**0.3193**	0.	0.0255
p-value (Pearson χ^2)	**0.3970**	**0.5338**	0.	**0.3083**

from creative employment, we get from Equations (16)-(18) that $\widehat{\alpha} = 1.675$, $\widehat{\beta} = 1.083$, $\widehat{\nu} = 11,329,658$, and $\widehat{n_{\min}} = 35,141.6$.

Table 4 presents goodness-of-fit comparisons as given by the Akaike Information Criterion and the Bayesian Information Criterion with other distributions. The distributions that are not rejected with a confidence level of $p = 0.05$ are shown in bold.

References

Abel, J. R., Dey, I., and Gabe, T. M. (2012). Productivity and the density of human capital. *Journal of Regional Science*, 52(4):562–586.

Acs, Z. J. (2002). *Innovation and the Growth of Cities*. Edward Elgar Publishing, Cheltenham, UK.

Agrawal, A. K. (2003). Innovation and the Growth of Cities (Book review). *Journal of Economic Geography*, 3(4):458–461.

Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution with Special Reference to Its Uses in Economics*. University of Cambridge, Department of Applied Economics, Monograph 5. Cambridge University Press, London.

Alves, L., Ribeiro, H., Lenzi, E., and Mendes, R. (2014). Empirical analysis on the connection between power-law distributions and allometries for urban indicators. *Physica A: Statistical Mechanics and its Applications*, 409(0):175 – 182.

Alves, L. G., Ribeiro, H. V., and Mendes, R. S. (2013a). Scaling laws in the dynamics of crime growth rate. *Physica A: Statistical Mechanics and its Applications*, 392(11):2672–2679.

Alves, L. G. A., Ribeiro, H. V., Lenzi, E. K., and Mendes, R. S. (2013b). Distance to the Scaling Law: A Useful Approach for Unveiling Relationships between Crime and Urban Metrics. *PLoS ONE*, 8(8):e69580.

Andersson, F., Burgess, S., and Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61(1):112–128.

Bacolod, M., Blum, B. S., and Strange, W. C. (2009). Skills in the city. *Journal of Urban Economics*, 65:136–153.

Barro, R. J. and Sala-i Martin, X. (2003). *Economic Growth*. MIT Press, 2nd edition.

Becker, G. S. (1964). *Human Capital: A Theoretical with Special Reference to Education*. New York: Columbia University Press.

Berry, B. J. L. and Okulicz-Kozaryn, A. (2012). The city size distribution debate: Resolution for US urban regions and megalopolitan areas. *Cities*, 29:S17–S23.

Berry, C. R. and Glaeser, E. L. (2005). The divergence of human capital levels across cities. *Papers in regional science*, 84(3):407–444.

Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science*, 340:1438.

Bettencourt, L. M. A., Lobo, J., and Strumsky, D. (2007). Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research Policy*, 36(1):107–120.

Bettencourt, L. M. A., Lobo, J., Strumsky, D., and West, G. B. (2010). Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE*, 5(11):e13541.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.*, 51:661–703.

Cooke, P., Asheim, B. T., Boschma, R., Martin, R., Schwartz, D., and Tödtling, F. (2011). *Handbook of Regional Innovation and Growth*. Edward Elgar Publishing.

Davis, D. R. and Dingel, J. I. (2014). The Comparative Advantage of Cities. NBER Working Papers 20602, National Bureau of Economic Research, Inc.

Dosi, G., Freeman, C., and Fabiani, S. (1994). The Process of Economic Development: Introducing Some Stylized Facts and Theories on Technologies, Firms and Institutions. *Industrial and Corporate Change*, 3(1):1–45.

Florida, R. (2004). *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books, New York.

Florida, R., Mellander, C., and Stolarick, K. (2008). Inside the black box of regional development–human capital, the creative class and tolerance. *Journal of Economic Geography*, 8(5):615–649.

Frank, S. A. (2009). The common patterns of nature. *Journal of Evolutionary Biology*, 22:1563–1585.

Frank, S. A. (2014). How to Read Probability Distributions as Statements about Process. *Entropy*, 16(11):6059–6098.

Frank, S. A. and Smith, E. (2011). A simple derivation and classification of common probability distributions based on information symmetry and measurement scale. *Journal of Evolutionary Biology*, 24:469–484.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly journal of economics*, 114:739–767.

Gabe, T. M. and Abel, J. R. (2012). Specialized knowledge and the geographic concentration of occupations. *Journal of Economic Geography*, 12(2):435–453.

Glaeser, E. L. (1999). Learning in cities. *Journal of Urban Economics*, 46(2):254–277.

Glaeser, E. L. (2011). *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Penguin Press, New York.

Glaeser, E. L. and Saiz, A. (2004). The Rise of the Skilled City. *Brookings-Wharton Papers on Urban Affairs*, 2004(1):47–105.

Gomez-Lievano, A., Youn, H., and Bettencourt, L. M. A. (2012). The statistics of urban scaling and their connection to zipf's law. *PLoS ONE*, 7(7):e40393.

Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic ouput. *Journal of Economic Growth*, 16:309–342.

Hausmann, R. and Rodrik, D. (2003). Economic development as self-discovery. *Journal of Development Economics*, 72:603–633.

Iammarino, S. (2011). Regional innovation and diversity. In Cooke, P., Asheim, B. T., Boschma, R., Martin, R., Schwartz, D., and Tödtling, F., editors, *Handbook of Regional Innovation and Growth*. Edward Elgar Publishing.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.

Jones, C. I. (1995). R&D-based models of economic growth. *Journal of Political Economy*, 103(4):759–784.

Jones, C. I. (2002). Sources of U.S. Economic Growth in a World of Ideas. *The American Economic Review*, 92(1):220–239.

Jones, C. I. and Romer, P. M. (2010). The new kaldor facts: Ideas, institutions, population, and human capital. *American Economic Journal: Macroeconomics*, 2:224–245.

Kremer, M. (1993). Population Growth and Technological Change: One Million B.C. to 1990. *The Quarterly Journal of Economics*, 108(3):681–716.

Limpert, E. and Stahel, W. A. (2011). Problems with Using the Normal Distribution–and Ways to Improve Quality and Efficiency of Data Analysis. *PLoS One*, 6(7):e21403.

Limpert, E., Stahel, W. A., and Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*, 51(5):341–352.

Lobo, J., Mellander, C., Stolarick, K., and Strumsky, D. (2014). The Inventive, the Educated and the Creative: How Do They Affect Metropolitan Productivity? *Industry and Innovation*, 21(2):155–177.

Lobo, J. and Strumsky, D. (2008). Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects. *Journal of Urban Economics*, 63:871–884.

Lucas Jr., R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

Mantovani, M. C., Ribeiro, H. V., Lenzi, E. K., Picoli, S., and Mendes, R. S. (2013). Engagement in the electoral processes: Scaling laws and the role of political positions. *Phys. Rev. E*, 88:024802.

Marx, M., Strumsky, D., and Fleming, L. (2009). Mobility, skills, and the michigan non-compete experiment. *Management Science*, 55(6):875–889.

Miguélez, E. and Moreno, R. (2013). What Attracts Knowledge Workers? The Role of Space and Social Networks. *Journal of Regional Science*, 3(0):1–28.

Mincer, J. (1958). Investment in human capital and personal income distribution. *The Journal of Political Economy*, LXVI(4):281–302.

Moretti, E. (2004). Chapter 51: Human capital externalities in cities. In Henderson, J. V. and Thisse, J.-F., editors, *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2243–2291. Elsevier.

Mulligan, C. B. and Sala-i Martin, X. (2000). Measuring Aggregate Human Capital. *Journal of Economic Growth*, 5(3):215–252.

Muneepeerakul, R., Lobo, J., Shutters, S. T., Gómez-Liévano, A., and Qubbaj, M. R. (2013). Urban Economies and Occupation Space: Can They Get "There" from "Here"? *PLoS ONE*, 8(9):e73676.

Nelson, R. R. and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change.* The Belknap Press of Harvard University Press, Cambridge.

Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Cont. Phys.*, 46(5):323–351.

Piore, M. J. and Sabel, C. F. (1984). *The Second Industrial Divide: Possibilities for Prosperity.* Basic books.

Rauch, J. E. (1993). Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics*, 34:380–400.

Redner, S. (1990). Random multiplicative processes: An elementary tutorial. *Am. J. Phys.*, 58(3):267–273.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *The Journal of Political Economy*, 94(5):1002–1037.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5):S71–S102.

Shearmur, R. (2012). Are cities the font of innovation? a critical review of the literature on cities and innovation. *Cities*, 29:S9–S18.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.

Simon, H. A. (1968). On judging the plausibility of theories. In Rootselaar, B. V. and Staal, J., editors, *Logic, Methodology and Philosophy of Science III*, volume 52 of *Studies in Logic and the Foundations of Mathematics*, pages 439–459. Elsevier.

Soo, K. T. (2005). Zipf's law for cities: a cross-country investigation. *Regional Science and Urban Economics*, 35:239–263.

Sornette, D. (2012). Probability Distributions in Complex Systems. In *Computational Complexity*, pages 2286–2300. Springer.

Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy.* Guilford Press.

Strumsky, D. and Thill, J.-C. (2013). Profiling U.S. Metropolitan Regions by Their Social Research Networks and Regional Economic Performance. *Journal of Regional Science*, 53(5):813–833.

Weil, D. N. (2012). *Economic Growth.* Prentice hall, New York, 3 edition.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Addison-Wesley Press, Cambridge.