Applying Distributional Approaches to Understand Patterns of Urban Differentiation

by

Andrés Gómez-Liévano

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved July 2014 by the
Graduate Supervisory Committee:

José Lobo, Co-Chair
Rachata Muneepeerakul, Co-Chair
Gerardo Chowell-Puente
Luís M.A. Bettencourt

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

Urban scaling analysis has introduced a new scientific paradigm to the study of cities. With it, the notions of *size*, *heterogeneity* and *structure* have taken a leading role. These notions are assumed to be behind the causes for why cities differ from one another, sometimes wildly. However, the mechanisms by which size, heterogeneity and structure shape the general statistical patterns that describe urban economic output are still unclear. Given the rapid rate of urbanization around the globe, we need precise and formal mathematical understandings of these matters. In this context, I perform in this dissertation probabilistic, distributional and computational explorations of (i) how the broadness, or narrowness, of the distribution of individual productivities within cities determines what and how we measure urban systemic output, (ii) how urban scaling may be expressed as a statistical statement when urban metrics display strong stochasticity, (iii) how the processes of aggregation constrain the variability of total urban output, and (iv) how the structure of urban skills diversification within cities induces a multiplicative process in the production of urban output.

Para mi madre y mi padre.

ACKNOWLEDGMENTS

My first thanks are to my sister, mother and father. Without their unconditional love, support and understanding this doctorate would not have been possible. To my uncle Rafa who has been a role model and friend, and has always showed interest in my research and ideas. At his home, he provided me with a comfortable environment to write the last parts of this dissertation. For this I must also thank Heidi for her welcoming attitude. And my best friends in Colombia, Santiago, Juan Ruy, Martin, and Juan Santiago, with whom my longing for a better society developed and pushed me to pursue a doctoral degree.

A person to whom I am, and will always be, indebted is my school mathematics teacher, Orlando Farigua. This dissertation can be said to have started during his classes, long ago.

I was very lucky that Geoffrey West, Luís Bettencourt and José Lobo gave me the opportunity of being part of the research group *Cities, Scaling and Sustainability* at the Santa Fe Institute. I met Geoffrey at a talk he gave at La Universidad de los Andes in 2009. He was the president of SFI at the time, and he was a distant hero for me for his work on Allometric Scaling in Biology. Fortunately, I was able to approach him after his talk, and this interaction changed my life and opened the door to a world of big questions and great collaborations. I thank him for all of that. Luís was instrumental in my subsequent doctoral training. He funded my visits to SFI and gave me his support, time, space and attention to pursue my research questions. Luís is a force of nature, and the momentum of his ideas has been an inspiration. I must offer special thanks to my advisor and co-chair, José Lobo. He is an exemplary person at all levels. Naturally kind, always generous, wise and inquisitive, I could not have wished for a better mentor and friend. I only hope that I will one day pass down what I have learned from him. I finally thank SFI for hosting me for many weeks of work that resulted in this document.

I have a special admiration and gratitude to my co-chair Rachata Muneepeerakul. Being close to him is intellectually energizing. He has generously been there to talk and to give feedback whenever I have required. Thanks to professor Jay Taylor, who always had his office door open to guide me through my research. To Kevin Stolarick, for also hosting a visit to the Martin Prosperity Institute. The collaboration and conversations with Kevin were a valuable interdisciplinary resource to me. I thank professors James Collins, Paige Warren, and Robert Boyd, who, through their courses and conversations, afforded me with an incredible window to the world of evolutionary biology. To Sander van der Leeuw, who supported me financially and introduced me to interesting ideas in sustainability science.

The staff of MCMSC was an incredible source of support in all respects. Especially Margaret and Sherry, who brought peace, tranquility and encouragement, when there seemed to be none. With Dawn and Kim, they are the columns that support the AMLSS program and I want to thank them from my heart. And then I must thank Georgianna Miller, from SHESC, for her patience with all my inquiries about academic logistics. I thank Carlos Castillo-Chavez for the financial support and the opportunity to teach mathematics to undergraduate students. Gerardo Chowell-Puente and Sherry Towers for generously providing me with financial support through exciting research until the end of my studies.

Close friendships were formed during these last years. Oscar represents the friend with whom I can speak and talk for hours about any topic. His friendship, advice, feedback and company are invaluable. Thanks to Mridul, Yannis and Arthur, who were my escape from academia, for their joy and companionship. To Dustin, for being a noble and kind friend. To Diego, for his enthusiasm for research and for his patience as a roommate. And to Emmanuel, Anarina, Maytee, Ilyssa, Arlene, John, Oyita, Miles, Kenny, Romie, Kamuela, Na, and Jordan, who, with their personality, good humor and authenticity, were a source of inspiration.

And finally, I am grateful to my companion in life, *my* María. She has not only been a source of emotional support and love, but she has also been my harshest and toughest critic, and I thank her for that.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION

The motivation behind the present dissertation is to understand what the obstacles to the economic development of human societies are. This is a question too big for any one to answer, much less in a single document. But the question, nonetheless, has been the motivation behind this work, and the work of many academics from a large array of disciplines, from medicine, architecture and engineering to history, sociology and political science, and most predominantly, economics. These disciplines offer different points of view for how to address questions about the obstacles to economic development. This dissertation rests on the view that these obstacles will be understood once we figure out how the socio-economic processes that create, occur in, and transform *cities* work.

In the last decade cities have become a hot topic of research. In part, the works of Jane Jacobs and Robert E. Lucas, Jr. (Jacobs, 1969; Lucas Jr., 1988) were fundamental to sway the researchers' attention from nations to cities. The reason, they argued, is because many of the proposed causes behind economic growth depend on effects that act locally (for an interesting counter argument read Polèse, 2005). For example, regarding the effects of skills and knowledge on economic growth, Lucas Jr. (1988, p. 37) writes: "The external effects [of human capital] have to do with the influences people have on the productivity of others, so the scope of such effects must have to do with the ways various groups of people *interact*" (emphasis mine). He continues, "we *know* from ordinary experience that there are group interactions that are central to individual productivity and that involve groups larger than the immediate family and smaller than the human race as a whole. Most of what we know we learn from other people" (Lucas Jr., 1988, p. 38). Finally, he concludes, "What can people be paying Manhattan or downtown Chicago rents *for*, if not for being near other people?” (Lucas Jr., 1988, p. 39).

But the recent focus on cities comes also from recent events. The year 2008 marked a transition point in human history in which city dwellers became a worldwide majority when compared to their rural counterparts, according to the United Nations[1]. In the World Economic and Social Survey 2013, the United Nations warned that new strategies are needed that address the impacts of urbanization to create a sustainable model of economic development (United Nations, 2013).

Indeed, the fact that half of the human beings now live in urban areas means that what they do in their lifetime will affect many more people than if they lived in rural areas. By living in a city, their actions will touch more dimensions of the world as a whole, simply because they will have access to more material and informational resources. A major question from this point of view is: how does the activity of single urban dwellers spread through society? Some of these human beings will experience drastic changes in their quality of life, income, education, and health. Some of these changes will create more overall wealth, while others will generate more poverty. Our lack of understanding for how this will happen is disturbing. In this state of affairs we need formal understandings of the processes of urbanization.

In this dissertation I develop some approaches that incite new questions to understand cities and open new avenues of research. I do this by building on the recent efforts coming out of the Santa Fe Institute in New Mexico that aim to develop a New Science of Cities (Bettencourt and West, 2010).

One of the distinctive aspects of these efforts is that they are motivated by the expectation that all cities share some *empirical* regularities regardless of time, place and culture (Bettencourt *et al.*, 2007a, 2013; Bettencourt, 2013; Ortman *et al.*, 2014). These regularities

[1] Cited in `http://www.unfpa.org/pds/urbanization.htm`

compel the development of a *scientific* theory of urban phenomena. This theory would explain these regularities based on a few principles about how cities function and develop, and would produce predictions that would be testable with data. Cities, according to this line of research, are different manifestations of the same phenomenon: human agglomerations giving rise to an ecology of social interactions that creates wealth and drives population growth, leaving a physical mark of infrastructure that feeds on itself to perpetuate human interaction.

One of the main regularities is an apparent simple empirical association between different urban metrics Y and population size N, in the form of a power-law relationship:

$$Y = Y_0\, N^{\beta}. \tag{1.1}$$

The novelty of this result does not lie in the fact that there is an association between urban metrics and population size. Also, it is not that the association is non-linear. The novelty in this discovery is that it is simple (Batty, 2008), and that the exponent β presents some unexpected regularities, especially when compared to its biological analogue (West *et al.*, 1997, 1999; West and Brown, 2005).

Bettencourt *et al.* (2007a) show that for measures of infrastructure, like number of gas stations, hospitals, length of roads, etc., the exponent is typically $\beta_{\text{infr.}} \approx 0.85$; For measures of socio-economic activity, like total Gross Domestic Product (GDP), crime rates, patenting rates, etc., $\beta_{\text{soc-ec.}} \approx 1.15$.

Such a relationship is reminiscent of other relationships in the history of Science that gave way to general scientific theories. Table 1.1 lists some few popular examples. This is a very short list among several similar scaling relationships that have revealed important aspects of how natural systems function.

A theoretical explanation of equation (1.1) was recently proposed by Bettencourt (2013). According to this explanation, equation (1.1) arises from the spatial mixing of agents, with

Table 1.1: Scaling Relationships That Gave Way to Deeper Understandings of Nature.

Quantities	**Scaling Law**	**Name**	**Theory**
Orbital period T and distance to the Sun r	$T = T_0\, r^{3/2}$	Kepler's third law	Newton's theory of planetary motion
Average radios of diffusion r and time t	$r = r_0\, t^{1/2}$	Law of diffusion	Einstein's theory of Brownian motion
Metabolic rate B and body mass M	$B = B_0\, M^{3/4}$	Kleiber's law	Metabolic Theory of Ecology

limited resources and subject to transportation costs, interacting through physical infrastructure (Bettencourt, 2013).

A note on the difference between this approach and others that have been proposed to understand quantitatively cities, coming mostly from the field of economics, is worth a few words. The importance of population size as a driver of per capita productivity has been implicit since the work of Adam Smith. The notion of the division of labor suggests, for example, that increases in productivity will arise from greater individual specialization, which is enabled by a larger population of workers. But the work in economics has been also an exploration about *which* factors, other than population size, affect productivity. In this sense, empirical works have been an enumeration activity, listing which factors play a role in urban productivity, and the theoretical works have followed to support such listings. Urban scaling analysis differs notably from these efforts. For example, one focus of analysis has been instead to explain the empirical value of β, *given that population size is the explanatory variable of aggregate productivity with the largest explanatory power* (Sveikauskas, 1975; Quigley, 1998; Ó hUallacháin, 1999; Bettencourt *et al.*, 2007a,b; Lobo and Strumsky, 2008; Bettencourt *et al.*, 2010): why $\beta_{\text{infr.}} \approx 0.85$ and why $\beta_{\text{soc-ec.}} \approx 1.15$? Perhaps by historical contingency, deriving the values of exponents in scaling relationships from theory, and understanding their meaning, has been typically in the interest of physicists, like the examples mentioned in table 1.1 (Schroeder, 1991; Whitfield, 2006). Indeed, Bettencourt (2013) has derived a value $\delta = 1/6$ for $\beta_{\text{infr.}} = 1 - \delta$ and why $\beta_{\text{soc-ec.}} = 1 + \delta$, and has linked this exponent δ to the efficient use of resources and energy within a transportation infrastructure that generates structured socio-economic interactions.

We hold in this dissertation that the science of cities must also explain the origin of the statistical distributions of urban metrics. This science should explain the origin of the fat-tailed distributions seen at the levels of individuals, such as in personal productivity (in crimes, inventivity, wages, etc.), and at the level of cities, such as the famous Zipfian distribution for population sizes (Zipf, 1949). More importantly, a statistical theory of cities should explain how all these different variables are related and affect others.

I will be referring mostly to the U.S. case, although the research presented in this dissertation is intended to apply to other countries as well (for example, chapter 3 will study the cases of Brazil, Colombia, and Mexico).

1.1 What Are Cities?

For our empirical studies of cities and urban systems we will use the geographical units that governmental agencies use, defined as *local labor markets*. And labor markets in this context are the geographical areas in which employers and employees can find a match. The idea behind such definitions is that cities are the places where individuals *live and work*. Thus, these definitions of cities also include suburban areas capturing the people that commute to the denser and more central places where most of the jobs are. In the U.S. they go under the name of Core Based Statistical Areas, but we will in general use the term "urban areas". The specific details of such definitions, which also depend on the country, will be given in later chapters.

However, since this dissertation goes beyond empirical studies and addresses theoretical questions as well, a few words about the the concept of *city* are worth discussing.

Despite the continued efforts to define cities as regional labor markets more accurately for socio-economic purposes, the establishment of these boundaries are still arbitrary, given that cities are open and there is a constant flux of people, not only from the vicinity, but from the whole urban system (usually the country to which they belong). Hence, how to define cities is not trivial, and the problem goes beyond decisions about where to put geographical boundaries (see Rozenfeld *et al.*, 2008, 2009; Arcaute *et al.*, 2013 for alternative definitions based on population densities).

Cities are usually classified by size and type, and thus we have megalopolis which refer to regions (or systems of cities), hypercities with over 20 million inhabitants, megacities with over 10 million, all the way to towns and villages. But how to define a city is not a trivial matter, and it mostly depends on the question and problem in hand (e.g., read Pickett *et al.*, 2011). For example, we have administrative boundaries (e.g. from one county to another), economical boundaries (e.g. metropolitan statistical areas), or ecological boundaries (defined by watersheds, mountain ranges, etc.). And different types such as pre/post-industrial cities or pre/post-World War II cities gives us a way to understand their spatial structure (Warren *et al.*, 2010).

Questions about the nature of cities go back to Plato and Aristotle, and probably even before (Portugali, 2000; Bettencourt *et al.*, 2009). Often, conceptual definitions of cities are expressed through comparisons to other phenomena (Bettencourt *et al.*, 2009 address quantitatively some of these comparisons). Thus, there is the view of cities as natural organisms (Graedel, 1999; Berry, 1964), as machines (Batty, 2012), as informational non-linear systems (Crosby, 1984; Portugali *et al.*, 2012), fractals (Frankhauser, 1994; Batty and Longley, 1994; Batty, 2005), as social networks (Jacobs, 1969, 1985; Glaeser, 2011), and more recently, as social reactors (Bettencourt, 2013). These analogies also come with their respective caveats about how different and unique cities are from these other phenomena. And yet, the debate about how to define cities seems to be endless (Berry and Okulicz-Kozaryn, 2012).

Given that our focus of analysis is on economic output, we pay special attention to one of the many aspects of cities: self-organized, physical systems constantly producing material and informational output. However, this dissertation *does not* offer a new definition of cities. It rather builds on already existing ideas of cities as inherently noisy, messy, finite, and complex collections of interacting elements (Batty, 2005; Bettencourt *et al.*, 2009; Storper *et al.*, 2012; Scott and Storper, 2014), to advance an already existing methodology in a context in which it has been seldom used. That is, the study of probability distributions (i.e., the distributional approach) of urban metrics beyond the conventional population size distributions, and the mechanisms they suggest.

1.2 What Advantages does a Distributional Approach Bring?

The guiding principles in economics of competitive markets, equilibrium of supply and demand, utility maximizing agents, and, in the particular case of urban economics, the spatial equilibrium hypothesis (Glaeser, 2008), are rich enough that they provide specific and testable predictions about the behavior of variables of interest. Hence, depending on the few additional assumptions and the question of interest, researchers in the field of urban economics have focused on studying, for example, the impact of human capital on productivity, or the positive and negative correlations between land rents, consumption and wages (Moretti, 2012). Among the many statistical methodologies that may be used to analyze data from cities, the most common, by far, have been econometrics and regression models.

"In its most canonical and popular form, a regression analysis becomes a "structural equation model" from which "causal effects" can be estimated" (Berk, 2010, p. 481). With this goal in mind, economists have advanced not only our knowledge of cities, but the tools of econometrics to analyze them and extract causal effects. With the advances in computational processing times, economists now have in their hands sophisticated and powerful statistical tools to test their ideas accurate and precisely. As explained by Berk (2010), though, the models to be tested must be "nearly right". And by this Berk means that the model must accurately represent reality. Which is, in turn, a problem. As a consequence, "[i]n the absence of a nearly right model, the many desirable statistical properties of a causal model can be badly compromised" (Berk, 2010, p. 482). Strategies to deal with problems of model misspecification exist, but in practice they often complicate the matters even more since additional new and untestable hypothesis are typically required.

Berk (2010) asserts that the solution to these problems is to broadly redefine regression analysis. Accordingly, he states that regression analysis, broadly understood, is "[to understand] as far as possible with the available data how the *conditional distribution of the response* [...] *varies across subpopulations* determined by the possible values of the predictor of predictors" (emphasis mine, Berk, 2010, p. 483). This definition, according to Berk, does not make any mention about addressing causal statements. Throughout this dissertation I will refer to this broad form of regression analysis as the *distributional approach*. Conventional regression analysis (e.g., econometrics), thus, consists of a particular step in the distributional approach in which we focus on the conditional mean. From this perspective, conventional regression analysis is a subset of a larger and more general distributional approach.

Distributional perspectives have been historically more common in the natural sciences, such as in physics (Sornette, 2006) and biology (Harte, 2011), than in the social sciences, such as the regional sciences. In economics, the study of probability distributions has been mostly limited to firm and city sizes, and financial price movements (see Gabaix, 2009 for an overview). Studying the intrinsic variability found in and across cities, and finding the commonalities among the differences, has started to be the subject of study of a few investigators (Bettencourt *et al.*, 2010; Gomez-Lievano *et al.*, 2012; Youn *et al.*, 2014; Bettencourt *et al.*, 2014; Alves *et al.*, 2013b,a, 2014; Storper *et al.*, 2012; Scott and Storper, 2014), but these efforts are still incipient, and we need theoretical developments that address this variability directly. Incorporating into our study of cities a distributional approach is necessary given that we not only want to understand the *tendencies* of the different variables, but also their *generative processes* (Sornette, 2012; Frank, 2009; Frank and Smith, 2011). And this is especially relevant in cities given their inherent heterogeneities. This dissertation is the first study that contributes to the regional sciences literature by studying probability distributions of urban aggregate output, generally understood, their meaning and relation to possible generative mechanisms.

The capacity of the distributional approach to suggest generative mechanisms is the main reason we adopt such methodology. By defining the full statistical distribution of our variables of interest, we are able to simulate the system under analysis (see, e.g., chapter 2, chapter 5, and appendix H), and have a more holistic analysis to the generated statistical patterns. Whereas the predictability in econometric analysis comes from being able to estimate average behaviors, from a distributional approach, predictability comes from the full statistical characterization of the phenomenon at hand.

The scientific process of understanding and discovery of the world does not discriminate among particular methodologies, as long as these methodologies enable the disproving of our hypothesis and theories (Deutsch, 2011). And some methodologies are more appropriate than others for answering particular questions. Hence, science advances the best when multiple methodologies are used to understand a phenomenon (Mayo, 1996).[2] There are certain instances in which the estimation of the conditional mean is enough. The following chapters in the dissertation will show, however, that a broader understanding of

[2]Mayo (1996)'s book, interestingly, advocates for the eclectic use of statistical strategies to identify errors and falsehoods in our hypothesis, but strongly criticizes Bayesian approaches.

cities emerges from applying a general regression analysis, as understood by Berk (2010), that not only focuses on the mean, but also on the deviations from it.

1.3 Overview of the Topics

1.3.1 Are Cities Finite Systems that Violate the Law of Large Numbers?

When studying economic productivity in cities, the first observation is that there are wide differences in productivities per capita across cities. This "stylized fact" still lacks a definite explanation, despite much progress in urban economics. In conventional regression analysis, the goal is to find the explanatory variables that best correlate with productivity per capita in cities and reduce the unexplained variance. But is there a different interpretation of this stylized fact? Is there a mechanism that would generate the patterns of productivity per capita that we see? Our answer is that there is at least one, and it consists of a violation of the law of large numbers when using per capita measures under some situations that we analyze.

Our study in chapter 2 stands as a counter argument to one recent criticism of the urban scaling approach, which was expressed as: "The impressive appearance of scaling . . . is largely an aggregation artifact, arising from looking at extensive (city-wide) variables rather than intensive (per-capita) ones." (Shalizi, 2011, p. 1). This criticism assumes that per capita transformations are always valid. We provide formulas for actually guiding this decision to use per capita measures.

This problem is a core part of our scientific understanding of cities, since we want to understand how systemic properties arise from the micro-components. We thus follow Uslaner (1976, p. 131)'s recommendation that "[p]er capita measures should not be shunned, but theoretically justified".

1.3.2 Urban Laws

The urban scaling law is really a statistical statement, and sometimes the deviations from it can be large. These deviations can even result in an apparent conflict with what scaling means. For example, how can equation (1.1) be valid if sometimes we observe many cities with $Y = 0$, as in homicides counts? This and other remarks highlight the need for a probabilistic statement of urban laws.

Bettencourt *et al.* (2013) note that "[a]n ultimate theory of cities should provide predictions about urban indicator statistics, including the expected value of deviations from the mean scaling prediction and the correlations in time and space of these deviations... A statistical theory of cities is necessary to eventually account for individual and collective variability in and across urban areas" (Bettencourt *et al.*, 2013, p. 6 & p. 19).

In chapter 3 we develop a statistical *description* of urban scaling at the aggregate level. Together with the scaling law expressed in equation (1.1), there exists another so-called law that describes the distribution of population sizes of cities for most urban systems: Zipf's law (Zipf, 1949; Mandelbrot, 1961; Gabaix, 1999, 2009; Batty, 2008; Giesen *et al.*, 2010; Malevergne *et al.*, 2011; Ioannides and Skouras, 2013). Two important results of this study are worth mentioning here: First, the exponent in the scaling law is a ratio between the Zipf exponents of the marginal distributions of population size N and the metric Y. And second, the *conditional* distribution of Y given N is approximately lognormal.

1.3.3 Human Capital

Innovation and knowledge creation have been identified in the literature as major determinants of urban economic growth and development, starting with the original "new growth" theories of Romer (1986, 1990) and Lucas Jr. (1988). Three empirical facts stand out: (1) Highly educated people concentrate disproportionately in larger cities (Rauch, 1993; Glaeser and Saiz, 2004; Bettencourt *et al.*, 2007b). (2) The population of places with high levels of human capital grows faster than places with low levels (Glaeser *et al.*, 1995; Acs, 2002). And (3), increases in human capital entail gains in regional productivity and innovation (Glaeser *et al.*, 1995; Rauch, 1993).

Whereas some of the literature has maintained that technologies are the embodiment of ideas and knowledge, part of the field in urban economics has shifted its attention instead to the sources of knowledge, that is, people (Jones, 1995; Bettencourt *et al.*, 2007b; Florida *et al.*, 2008; Jones and Romer, 2010). However, although the early research studies on the effects of human capital on productivity improved our understanding of cities, their scope was too narrow, because they limited their definition of human capital only to the level of educational attainment.

The recognition that knowledge is about the accumulation of ideas, and that ideas are ultimately originated in individuals (Jones, 1995; Jones and Romer, 2010), shifted the emphasis towards skills and the notion of "tacit knowledge". In this context, spatial proximity matters. And it is population size, the diversity of skills found in individuals (i.e., the tacit knowledge that people accumulate through education and experience), and what they *actually do* with them, that set a more adequate basis on which to study the economic performance of cities. In this line, the work of Florida (2004, 2011); Florida *et al.* (2008) redefined what is understood by "human capital" with their definition of creative class employment. As Florida notes, "[w]hen talented and creative people come together, the multiplying effect is exponential; the end result is much more than the sum of the parts. Clustering makes each of us more productive–and our collective creativity and economic wealth grow accordingly" (Florida, 2011, p. 193).

But is the accumulation of "talented and creative people" constrained in some way? Answering this question is important since it is assumed that constraints in the creative and inventive activities in urban areas would hamper economic growth. We answer this question in chapter 4.

1.3.4 The Structure of Diversification

In urban studies there has been a long and heated debate about whether urban economies benefit more from specialization or from diversification[3] in their composition of firms and individuals (see Beaudry and Schiffauerova, 2009; Polèse, 2013 for longer discussions about this topic). Specialization occurs when benefits arise from producers and consumers of similar industries agglomerate in the same place, resulting in what are called Marshall-Arrow-Romer (MAR), or localization, externalities. Conversely, diversification occurs if there are benefits that producers and consumers can exploit from the variety that cities can provide; these are traditionally called Jacobs, or urbanization, externalities (Jacobs, 1969; Glaeser *et al.*, 1992). This topic is part of a more fundamental question in urban economics of why people and firms agglomerate in cities in the first place despite incurring in associated costs of congestion and higher land rents. By now, the empirical literature has in general shifted to support the Jacobs' type of benefits to diversity (Quigley, 1998). The models addressing the problem of agglomeration economies emphasize three mechanisms (Duranton and Puga, 2004; Rosenthal and Strange, 2004, 2006; Puga, 2010): *sharing* (e.g., of goods, facilities, risk), *matching* (e.g., for job seekers and employers), and *learning* (e.g., as in knowledge spillovers).

Much of the corresponding theory on economic development has been done in the framework of Cobb-Douglas production functions that only use aggregate levels of physical and human capital as inputs. This conceptualization does not address the question about the specific mechanisms which transform diversity into productive output (see, however, Duranton and Puga, 2004). In fact, the neglect of diversity is already implicit in the notion

[3] I will use the words "diversification" and "diversity" interchangeably.

of "substitution", which is one of the basic assumptions behind Cobb-Douglas production functions. Ironically, the recognition that not only the *extent* (captured by population size), but also the *structure* of diversity in cities are important to economic productivity can be traced back to some of the earliest works in urban economics. Jacobs (1969), for example, emphasized that economies not only *expand* by doing more of the same, but also *develop* by adding new kinds of work. Similarly, for Sveikauskas (1975, p. 394), "[c]reativity, or the successful adaptation to change, can be thought of as the rearrangement and recombination of hitherto separate elements." And later on, Lucas Jr. (1988, p. 35) stated that "a successful theory of development... has to involve more than aggregative modeling".

In the context of economic growth at the level of countries and the portfolio of products that countries export, Hidalgo *et al.* (2007); Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011) have developed a framework to think about economic growth as a process of economic diversification. They propose a production function that transforms the complexity of "capabilities" (i.e., skills and tacit knowledge) that countries possess, into the explicit diversity of goods that nations produce, thus going beyond explanations that rely only on the aggregated levels of physical and human capital.

Brummitt *et al.* (2012) show, however, that standard measures of occupational diversity in the US such as the Herfindahl-Hirschman index or Shannon's entropy have a weak statistical predictive power over urban output, once urban scaling with population size has been controlled for. This suggests the possibility that it is the structure of interdependencies between occupations that has some explanatory power over urban economic performance (Muneepeerakul *et al.*, 2013). Recent studies are starting to shed evidence that a structural view of the urban productivity is needed (Neffke and Henning, 2013; Neffke *et al.*, 2013). This approach is in contrast with the traditional regression-based studies that have in the last years increased in size and sophistication. In chapter 5 our motivation is to know whether this can be formalized in a mathematical and computational model. Given the empirical patterns we observe, we want our model to answer the following questions: after controlling for population size, is the remaining unexplained lognormal variation in the levels of productive output, e.g., in Gross Metropolitan Product or Patenting, a result of *differences in the structure of urban diversification?* Our answer, from the point of view of the model we propose, is yes.

Chapter 2

A DISTRIBUTIONAL APPROACH TO URBAN PRODUCTIVITY

In this chapter we introduce some of the motivations to use distributional approaches to deepening our understanding of cities. To do this, we start with one of the fundamental questions in urban economics: What explains the wide productivity differentials across urban areas? We present a parsimonious null model of an urban system that reproduces some qualitative features of real world data. The model highlights (1) the importance of considering the internal heterogeneity existing in cities using probability theory, and (2) the importance of examining how aggregate measures of productivity change with population size as a way to understanding the underlying generative processes.[1]

2.1 Introduction

Figure 2.1 plots the number of patent applications per capita in a cross section of the U.S. Metropolitan and Micropolitan Statistical Areas in the year 2000, against population size. As can be seen from the figure, there is significant variation in productivity across urban areas. These different levels of patent production are, presumably, a good reflection of the differences in the average productivity of the individuals living in those places. The question is thus: Why are some urban areas more productive than others?

We will not provide a direct answer to this question, which has been one of the central subjects of study of the urban economics discipline (Polèse, 2013). We address a related question instead: What interpretations should we give to our productivity measures? In other words, How should we interpret what we see in fig. 2.1, and what methodology

[1]This work was done in collaboration with Professors Vladislav Vysotsky and José Lobo.

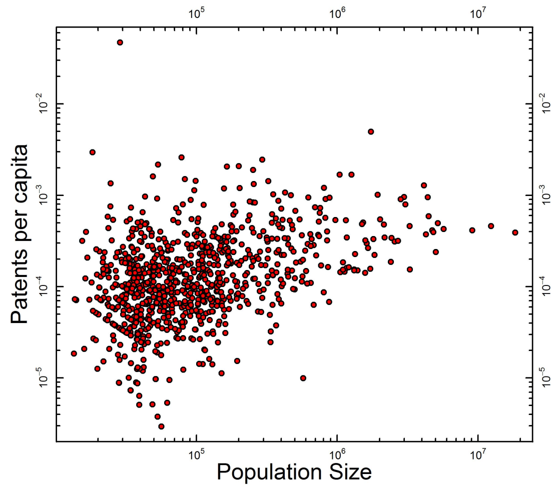
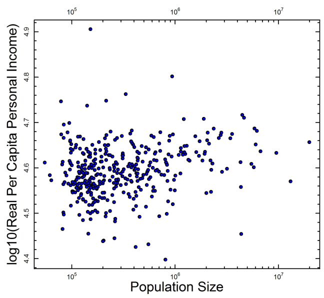


Figure 2.1: Patents Per Capita in U.S. Metropolitan and Micropolitan Statistical Areas (CBSAs) Against Population Size in the Year 2000. Urban Areas with Zero Patent Applications Are Not Shown, and as a Consequence, Only 895 Dots Are Shown out of 938. Source: U.S. Patent and Trademark Office.

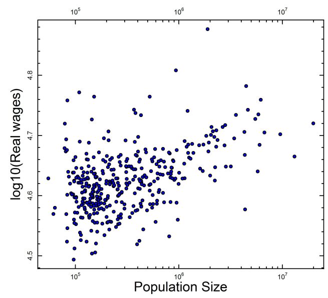
should we use? We are going to show a set of mathematical and computational results which challenge the prevailing view that the large productivity differentials come from a few, yet to be discovered, factors of production. Rather, we will argue that the reason behind the differences in productivity in fig. 2.1 may be that urban processes are intrinsically stochastic. If this is the case, our theories should reveal the mechanisms behind those processes.

We will show that if the goal is to understand the mechanisms underlying the differences in productivity of urban areas a more adequate way of looking at the data is through aggregate measures of productivity, as opposed to per capita measures. That aggregate measures must be studied under the lenses of scaling analysis, and that their fluctuations, given population size, must be objects of a distributional analysis.

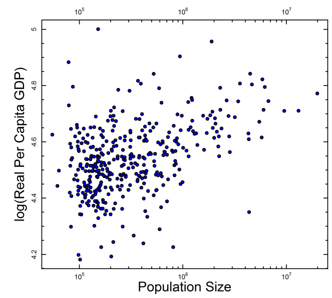
The qualitative features of fig. 2.1 are not unique to patents per capita, and also appear in other socio-economic measures of productivity per capita. Figure 2.2 plots real personal income (fig. 2.2a), real wages (fig. 2.2b), and real GDP per capita (fig. 2.2c), for the U.S.



(a) Price-Adjusted Estimates of Real Per Capita Personal Income, Chained 2008 Dollars.



(b) Real Wages (See appendix A.1.3 for the Estimation), Chained 2008 Dollars.



(c) Per Capita Real GDP, Chained 2005 Dollars.

Figure 2.2: Cross Section of Different Measures of Productivity in the 381 U.S. Metropolitan Statistical Areas in the Year 2012. Source: U.S. Bureau of Economic Analysis.

381 Metropolitan Statistical Areas (see appendix A for details about the sources of the data). These figures display the same features: large variance in the vertical dimension and a slight, but statistically significant, positive dependence of the productivities per capita with population size[2]. Although the topic of the present chapter is more theoretical than empirical, we will keep patents as our reference measure of productivity because it is an unambiguous measure of productivity and because it is also available for Micropolitan Statistical Areas. Chapter 4 will explain in more detail the construction of this database.

In principle, measures of productivity per capita have the advantage of correcting for size effects. The fact that after controlling for population size there is still a large variation left unexplained in urban productivity, such as in fig. 2.1, has motivated a large body of research in urban economics. Guided by economic principles, researchers have studied a

[2]Personal income is the income received by all persons from all sources. One must be careful in interpreting personal income as a measure of productivity, however, since it also includes non-local sources such as stock investments, government transfers, rents, etc.. See `http://www.bea.gov/scb/pdf/2013/08%20August/0813_regional_price_parities.pdf` for details about the adjustments to control for prices level differences across regions.

diversity of causes behind productivity differentials. In doing so, they have advanced and developed sophisticated econometric models to test the effect of different factors. Thus, productivity per capita has been regressed against levels of education, infrastructure, industrial concentration or diversification, crime, and many others. Despite much progress in the identification of factors best associated with greater productivity (see Puga, 2010), here we will propose a different approach for how to view productivity in cities (see table 2.1).

Table 2.1: Two Different, Although Not Incompatible, Causes and Interpretations About the Large Variance Displayed in fig. 2.1, and the Corresponding Methodology That Can Be Used to Understand it.

Interpretation		**Methodology**
1. Few, specific, yet to be identified factors	$\Longrightarrow$	Econometrics
2. Inherent stochasticity in the generative mechanisms	$\Longrightarrow$	Distributional Analysis

Beyond typical issues about the independence or normality of the errors, or about the problems of endogeneity of the regressors, conventional econometric models implicitly assume that:

1. Per capita measures offer valid and representative information of the individuals living in a city.

2. There is an *exact* and *deterministic* relation between productivity per capita and some (yet to be discovered) factors that will reduce the error terms in the regression equations to zero at best, or to a weak white noise, stemming from measurement errors, at worst.

We contend that the violation of these two assumptions may be an important characteristic of cities and urban systems. Namely, we should consider the possibility that per capita metrics of productivity may be inadequate measures of the productivity of individuals and that the deviations may carry important information not only about unobserved factors, but about the *processes* that generated the data. Given the large variances observed, we should expect these processes to be highly stochastic.

The null model presented in the following sections will help prove the point that the conventional assumptions associated with econometric models might not hold. In doing so, we will show that (1) it is best to leave measures of productivity in their aggregate version (as opposed to taking a per capita transformation), (2) that population size is an important, if not the main, explanatory variable that must be included when explaining productivity, and (3) that the distribution of the aggregate measures conditioned on size can be informative about the internal workings of cities.

2.2 Aggregate or Per Capita Metrics?

Our understanding of the functioning of cities depends on our understanding of what happens to the individuals living in them. Information at the level of the individuals, however, is often unavailable and only aggregate and coarse grained measures at the level of the whole city are at reach. As a consequence, information about the individuals in a city is typically inferred through per capita metrics, which come from dividing the total aggregate measure by the population size. Ideally, per capita metrics will represent the average behavior of individuals in the city.

However, as our ability to probe cities at finer scales grows we increasingly uncover broad heterogeneities within them. Indeed, recent studies show that inequality is growing in time (Piketty and Saez, 2003; Banerjee and Yakovenko, 2010), and is worse for bigger cities than in small ones (Berube, 2014). More and more, the idea of an "average individual" has been losing its meaning. We do not enter the debate around the topic of inequality in cities (e.g. Glaeser *et al.*, 2009), but it does serve us as an additional reason to study the consequences of broad inequalities and population size on the appropriateness of per capita, as opposed to aggregate, measures.

Per capita measures are the measurement *par excellence* of urban socio-economic productivity, be it for wealth, invention or crime. But the decision to use a per capita measure to analyze systemic properties is actually a non-trivial problem that has to be decided based on formal arguments. Although the problem of what to measure about a system depends on the analytical context and the questions of interest, here we emphasize that the statistical description of urban systems, and social systems more generally, renders per capita measures inappropriate in many contexts.

Note that we are not arguing against the use of per capita measures as purely *descriptive* tools. If, for instance, the total wealth of a metropolitan area is Y, and its population is N, both measured to a high degree of accuracy, there is nothing to debate about the fact that Y/N *really is* the corresponding metropolitan area's empirical per capita wealth. What is debatable is whether this value is representative of the mean of the generative process that created the individual components of Y. As we will demonstrate shortly, there are situations that make the use of per capita measures problematic. In these situations, for example, the same generative process with the same mean will generate a larger productivity per capita in large cities than in smaller ones, revealing an apparent paradox. Its resolution will be the heart of our argument for the use of scaling analysis and a distributional approach to understanding cities.

The problem with per capita measures is commonly expressed as a methodological one. Some concerns were expressed and analyzed early on, for example, by Uslaner (1976), who demonstrated that the nonlinearities implicit in per capita transformations can induce misleading interpretations of the data. Part of the problem is that individual-level properties are often dependent on the size of the system. Therefore the ratio of two system-wide measures, each of which aggregates many size-dependent effects (e.g., dividing GDP over population size), may not stand as a meaningful indicator. In other words, the feedback between the behavior of the parts and the whole imposes constraints on the type of intensive measures that can be defined (see Katz, 2006; Bettencourt *et al.*, 2010). Here we show that even in simpler systems where this feedback is not present, per capita measures still stand as inadequate indicators of individual-level properties.

We explore this problem in the light of probability theory. A per capita measure is an estimate of the mean expected value of an individual-level random variable. In what follows, we use the random variable X to represent the productivity of an individual. And we will denote the population size and the aggregate productivity of an urban area by N and $Y_N \equiv \sum_{i=1}^{N} X_i$, respectively. Intuitively, the validity of per capita measures is guaranteed by the Law of Large Numbers (LLN). The (weak) LLN states the following:

Theorem (Weak Law of Large Numbers)**.** *Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed (i.i.d.) random variables with mean $\mathrm{E}[X_1] \equiv \mu < \infty$, and let $Y_N = \sum_{i=1}^{N} X_i$. Then,*

$$\Pr\left\{\left|\frac{Y_N}{N} - \mu\right| \geq \varepsilon\right\} \overset{N\to\infty}{\longrightarrow} 0, \tag{2.1}$$

for all $\varepsilon > 0$.

In other words, as the number of terms N increases, the probability that the sample average Y_N/N differs from the true average value μ by an amount ε tends to zero.

The two conditions for the LLN to hold are that the mean exists and is finite, and that the number of terms N in the sum Y_N be infinite. Since the latter condition is never met in practice (all systems have finite number of components), one must define an estimation interval to approximate μ using Y_N/N. This determines a region of convergence, i.e., a minimum number of terms after which it matters not whether N reaches infinity. This region of convergence, in turn, is determined by the broadness or narrowness of the distribution of X_1.

We present in the next section the simplest example of the effects of violating the LLN, which I will refer to as the *Lévy Case*. This will set a point of reference for how to think about our null model.

2.2.1 The Lévy Case

The first condition of the LLN is that $\mathrm{E}[X]$ must be finite. This condition is violated, for example, if the right tail of the probability density function (pdf) of a random variable X is Paretian with an exponent $\tau \leq 1$, meaning that the pdf decays as a power-law

$$p_X(x) \propto x^{-\tau-1} \quad \text{for } x \geq x_0 > 0. \tag{2.2}$$

When $\tau \leq 1$, the integral $\int x p_X(x) \mathrm{d}x$ diverges, which formally shows that the variable X lacks a well-defined mean. As a consequence, taking the arithmetic average of a corresponding sample $X_1, \ldots, X_N$ of i.i.d. random variables from this distribution is not justified, *no matter how large N is*. This is because the Law of Large Number will never hold in this particular case. As N increases, the variable Y_N/N displays larger and larger fluctuations, instead of converging to a single value.

In spite of this behavior, the lack of a mean in equation (2.2) (for $\tau < 1$) has an interesting consequence over Y_N. Namely, that

$$\widehat{\mathrm{E}[Y_N]} \propto N^{1/\tau}, \tag{2.3}$$

where the hat symbol $\widehat{\ }$ means the estimated value from a finite number of realizations of the random variable Y_N. We refer the reader to appendix C for a proof.

There is a vast literature in probability and statistics addressing these and other properties of heavy-tailed probability distributions (see Embrechts *et al.*, 1997; Kleiber and Kotz, 2003; Newman, 2005; Sornette, 2006, 2012 and references therein). We believe there are useful insights from this literature that can be applied in the context of urban economics and regional science.

In the urban context, as we shall see, even for variables with well-defined *means and variances*, problems associated with taking arithmetic averages may exist, stemming from the characteristic finiteness of cities. Thus, both the broad intravariability that individuals in social systems often display (see references in Andriani and McKelvey, 2007) and the finiteness of cities will prevent meaningful interpretations of per capita quantities.

Let us introduce the null model to show how cities may in fact be violating the LLN.

2.3 A Simple Null Model of an Urban System

Model: Let a single city be defined as the collection of N i.i.d. non-negative random variables X_i, $i = 1, \ldots, N$, drawn from a lognormal distribution $\mathscr{LN}(a, \sigma^2)$ with probability density function

$$p_X(x; a, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(\ln x - a)^2}{2\sigma^2}}, \tag{2.4}$$

with mean $\mathrm{E}[X] = \mu \equiv \int x p_X(x) \mathrm{d}x = \mathrm{e}^{a+\sigma^2/2}$. The variables X will represent the productivities of individuals in a city, but could represent the productivity at a higher level of organization such as a firm.

We generate all the productivities of all individuals for m cities according to equation (2.4). For city $k \in \{1, \ldots, m\}$ with total population N_k, by definition, adding the individual variables results in the aggregate output level $Y_{N_k,k}$:

$$Y_{N_k,k} = \sum_{i=1}^{N_k} X_{i,k}. \tag{2.5}$$

The subscript k will in general be omitted, unless explicitly needed.

Figure 2.3 plots again fig. 2.1 (bottom-right panel), but shows additionally the simulation of 895 synthetic cities (top row panels) from this null model. In the simulated data, each dot is a city, and each city has a population N taken from the real U.S. Metropolitan and Micropolitan Statistical Areas. Hence, both the real and synthetic data consist

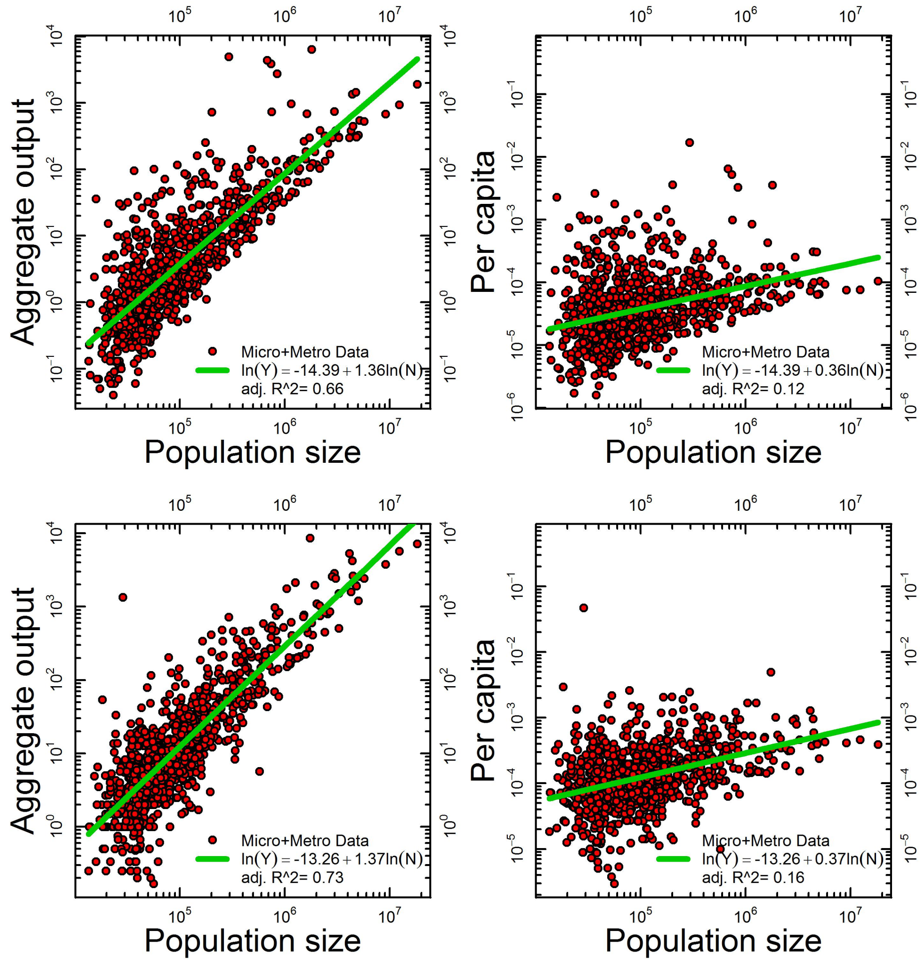


Figure 2.3: Top Row: Synthetic Simulation of Aggregate Output Y for Population Sizes N Taken from the Real U.S. Metropolitan and Micropolitan Statistical Areas in the Year 2000. For Each City, the Aggregate Output is $Y = \sum_{i=1}^{N} X_i$, Where $X_1, X_2, \ldots, X_N$ are i.i.d. Random Variables Lognormally Distributed $\mathscr{L}\mathscr{N}(a, \sigma^2)$. The Parameters a and σ Have Neen Set so that $\mu = e^{a+\sigma^2/2} = \sum Y_k / \sum N_k$, Where Y_k and N_k Represent the Real Total Number of Patents and Population of the k-th Urban Area. And We Set $\sigma = 6.0$, Which, According to fig. 2.5, Results in a Scaling Exponent of $\hat{\beta} \approx 1.37$. Bottom Row: Real Data, in Which the Output Y Represents the Number of Patent Applications in the Year 2000 Assigned to Each Micropolitan and Metropolitan Area.

of the same number of cities with the same population sizes, but they differ in their aggregate productivities. The synthetic total output $Y_N = \sum_{i=1}^{N} X_i$ was computed by simulating the whole collection of X_i, for each city. All variables X_i for *all* cities were generated from the same lognormal distribution with parameters $\sigma = 6.0$ and a such that $\mathrm{E}[X] = \mu = \mathrm{e}^{a+\sigma^2/2} = \sum Y_k / \sum N_k$ (i.e., the patents per capita across all urban areas), where Y_k and N_k represent the real total number of patents and population of the k-th urban area. The panels to the left and right represent the same information. The only difference is that to the left total aggregate values are shown and to the right the per capital values.

The parameter σ in our null model controls both the slope of the productivity per capita with population size and the overall spread around this average trend (the parameter a controls the intercept). Figure 2.4 shows this effect for six different values of σ. The black dashed line is $\mathrm{E}[X]$, the expected value of per capita number of patents. Note that $\mathrm{E}[X]$ does not have a dependence on N, whereas Y_N/N does (purple solid line), despite the fact that $\mathrm{E}[Y_N/N] = \mathrm{E}[X]$. This is reminiscent of the Lévy Case presented in section 2.2.1, except this time not only is $\mathrm{E}[X]$ finite, but also all the other moments $\mathrm{E}[X^n]$.

Figure 2.5 plots the effect of σ on the exponent $\widehat{\beta}$ of the regression between the synthetic $\ln(Y)$'s and their corresponding $\ln(N)$. As the value of σ increases and the tails of the lognormal become increasingly heavier, the value of β also increases, which means that the total output Y_N scales more and more steeply with larger population sizes. The similarity of the simulated data (top-right panel) to the real data (bottom-right panel) in fig. 2.3 was thus generated from this simple null model. By adjusting the parameter σ so that β equals the real scaling exponent: for $\widehat{\beta} = 1.37$, we set $\sigma = 6.0$. It is observed that the exponent $\beta \approx 1.16$ reported by the recent literature in urban scaling (see Bettencourt *et al.*, 2007a; Bettencourt, 2013) is reached for $\sigma \approx 4.5$.

The possibility that these empirical scaling exponents may be due in part to the fat tails of the distribution of individual productivities has a direct effect on our use of per capita metrics of productivity. Figure 2.6 plots the convergence of Y_N/N as N increases for the specific case of $\sigma = 4.5$ and $\mathrm{E}[X] = \mu = 1$. The quantiles $Y_N^{(p)}$ for $p = 0.05$ and $p = 0.95$,

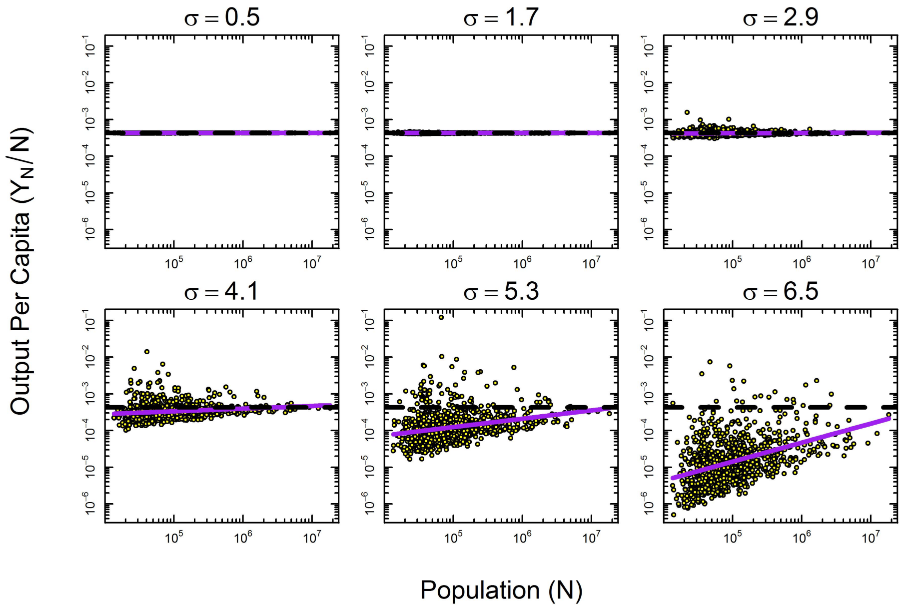


Figure 2.4: Effect of σ on Urban Productivity Per Capita in Null Model. In this Simulations, the Populations Are the Real Populations of U.S. Metropolitan and Micropolitan Statistical Areas in the Year 2000. The Parameter a is Changed so that $\mathrm{E}[X]$ is Fixed for All Cities and is Equal to the Productivity Per Capita Over All Urban Areas (Black Dashed Line). The Increasing Productivity Per Capita with Increasing Population Size is Shown Using an Ordinary Least Square Regression Line of $\ln(Y_N/N)$ Against $\ln(N)$ (Purple Solid Line).

defined such that

$$p = \Pr\left\{Y_N \leq Y_N^{(p)}\right\}, \tag{2.6}$$

scale as $Y_N^{(0.05)} \approx \mu - \exp(a_0 + a_1 \ln N)$ and $Y_N^{(0.95)} \approx \mu + \exp(b_0 + b_1 \ln N)$. Ordinary Least Squares estimations of these parameters yield $a_0 = 3.123 \pm 0.027$, $a_1 = -0.263 \pm 0.001$, $b_0 = 4.345 \pm 0.179$, $b_1 = -0.305 \pm 0.009$. With this we can calculate that the lower quantile will reach the threshold $\mu - \varepsilon = 0.9$ at around $N \approx 1 \times 10^9$, whereas the upper quantile will reach the threshold $\mu + \varepsilon = 1.1$ at around $N \approx 3 \times 10^9$. In other words, we can be 90% confident that our per capita estimation will be within an interval $[\mu - 0.1, \mu + 0.1]$ for populations in the order of *billions* (see fig. 2.6).

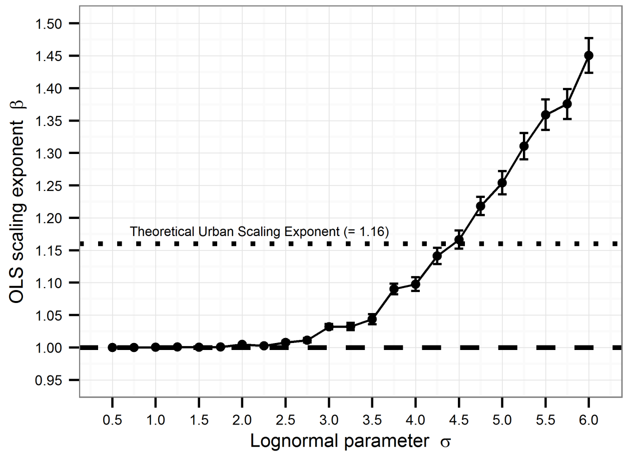


Figure 2.5: Departure from the Law of Large Numbers for a Lognormal Distribution as the Intra-Variability of the Components Within Cities Increases. Each Point in the Plot Shows the OLS Estimations of the Scaling Exponent β Between Aggregate Output Y and Population Size N, for 895 Simulated Cities, for a Given Value of the σ-Parameters of the Individuals' Lognormal. As σ Increases, the Scaling Exponent also Increases, Showing Increasing Violations of the Law of Large Numbers. The Error Bars Represent the Standard Deviation of the Estimate of β from the OLS Method.

This null model presents a naive view of a city, and its assumptions are unlikely to hold in reality. On the one hand, the distribution of individuals' productivity is unlikely to be as fat-tailed as a lognormal with a parameter $\sigma = 4.5$, much less $\sigma = 6.0$, since this would represent a distribution with gini coefficients larger than 99%, whereas data suggest that real gini is around 60%. On the other hand, the distribution itself is likely to be different across cities. The aim of the model, however, is not to model reality accurately, but rather to motivate a different approach to studying productivity differentials across cities.

In the next sections we relax the assumption of the lognormal distribution in the null model. And we present analytical results that establish how large must be N, in relation to how fat-tailed the distribution of $X_1, \ldots, X_N$ is, given a confidence level in our use of per capita measures.

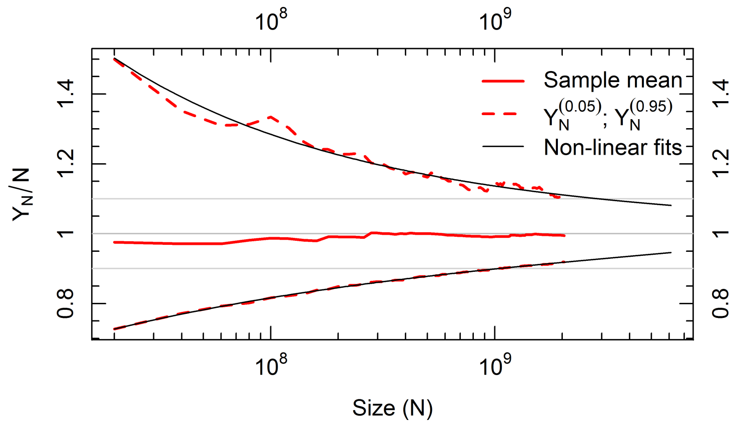


Figure 2.6: Convergence to the Law of Large Numbers for a Lognormal $\mathscr{LN}(-4.5^2/2, 4.5^2)$. Here, 1000 Partial Sums Y_N Were Simulated. In the Plot we Show the Convergence of Y_N/N to $\mu = 1$. The Red Dashed Lines Represent the 5% and 95% Quantiles, such that 90% of the 1000 Simulations Lie Between those Two Lines. The Red Solid Line is the Arithmetic Average of All the 1000 Values of Y_N/N at Each Point N. The Convergence is Very Slow.

2.4 Mathematical Statement of the Problem Regarding the Validity of Per Capita Measures

Suppose we observe a city of (finite) population size N and total aggregate output Y_N (e.g., total wealth or total number of patents produced). By definition, $Y_N = \sum_{i=1}^{N} X_i$, where $X_1, X_2, \ldots, X_N$ are the unobserved outputs of all the individuals living in the city. Assume the variables of the individuals are a non-negative i.i.d. sequence of size N from an *unknown distribution* with mean $\mathrm{E}[X] = \mu$. The problem is to find the constraints on the distribution of the individuals' variables X_i such that taking a per capita transformation of Y_N is "justified". By justified we mean that there is a high probability, at least $1-\alpha$, that $\mu - \varepsilon \leq Y_N/N \leq \mu + \varepsilon$, for a given small $\varepsilon > 0$. Mathematically,

$$\Pr\left\{\left|\frac{Y_N}{N} - \mu\right| \leq \varepsilon\right\} \geq 1 - \alpha. \tag{2.7}$$

Note that equation (2.7) imposes some constraints on the moments of the distribution of X.[3]

The question is: Given N, Y_N, $1 > \alpha > 0$ and $\varepsilon > 0$, what are the constraints on the distribution of X_i's under the condition in Equation (2.7)? Conversely, if we have knowledge about the distribution of X (e.g., through some of its moments), what is the minimum population size N^*_{min} above which inequality (2.7) holds?

The convergence of Y_N/N to μ is guaranteed by the LLN for 'large' N. Different distributions have different rates of convergence, however, and some can be very slow (e.g., fig. 2.6). As a consequence, how 'large' N should be depends critically on the lightness or heaviness of the tail of the distribution of X. When X_i's are *lightly-tailed* they do not differ from each other substantially, they each contribute approximately equally to the sum Y_N, and the convergence is fast. In contrast, when they are *heavy-tailed* differences between the X_i's are large, the maximum $M_N = \max\{X_1, \ldots, X_N\}$ among them can account for a large fraction of the total sum, and the convergence is slow.[4]

Since cities typically have populations that are considered large (on the thousands and up to tens of millions) a naive (and incorrect) answer to the question about the minimum N above which we can feel confident of taking a per capita estimation, is to assume that the variance v of y_i is finite and then invoke the Central Limit Theorem (CLT). We would then

[3] In appendix B we state an interesting question that arises from the analysis in this chapter. Namely, what would be Markov's inequality analogue for a random variable X in the situation when we have an estimation of $\mathrm{E}[X]$ only? A similar question was posed by Saw *et al.* (1984) for Chebyshev's inequality.

[4] This is specifically captured by the *subexponential* family of distributions. See Embrechts *et al.* (1997).

be inclined to say that $Y_N/N \sim \mathscr{N}(\mu, v/N)$, and thus

$$\begin{aligned}
\Pr\left\{\left|\frac{Y_N}{N}-\mu\right| \leq \varepsilon\right\} &= \Pr\left\{\frac{\frac{Y_N}{N}-\mu}{\sqrt{v/N}} \leq \sqrt{\frac{N}{v}}\varepsilon\right\} \\
&\quad - \Pr\left\{\frac{\frac{Y_N}{N}-\mu}{\sqrt{v/N}} \leq -\sqrt{\frac{N}{v}}\varepsilon\right\}, \\
&\approx 2\Phi\left(\sqrt{\frac{N}{v}}\varepsilon\right) - 1, \qquad (2.8)
\end{aligned}$$

where $\Phi(z)$ is the standard normal cumulative distribution function (cdf). For $\alpha = 0.01$ and $\varepsilon = 0.1\sqrt{v}$, inequality (2.7) is met for $N \geq 385$. Assuming this argument is correct, per capita measures are fully justified, since cities typically have more than 385 individuals. The problem is, however, that the convergence of the CLT represented by the approximation $\Pr\left\{(Y_N/N-\mu)\sqrt{N/v} \leq z\right\} \approx \Phi(z)$ may not be uniform in z, and even if it is, it may not be fast enough. Uniform convergence in z is met when the third moment of y_i's exists and is finite[5]. In this case, the rate of convergence is of the order $O(1/\sqrt{N})$. But for large N the term $\Phi(\sqrt{N/v}\varepsilon) \sim N^{-1/2}\mathrm{e}^{-N\varepsilon^2/(2v)}$ and, as a consequence, the Gaussian approximation can largely underestimate the probability of large deviations of Y_N/N from the per capita mean μ.

2.5 Analytical Results

The previous sections presented some numerical examples of how the interpretations of per capita quantities can provide misleading information about the mean of a distribution, and how violations of the LLN create an interesting systemic behavior whereby productivity scales superlinearly with population size. Here we present some analytical results to understand the scaling behavior, and the interplay between population size and the broadness of the distribution under a given confidence our per capita estimations.

[5]See *Berry-Esseen Theorem*, Shiryaev, 1995

2.5.1 Relationship Between Intravariability and Scaling

Intuitively, the behavior displayed in figs. 2.3, 2.4, and 2.5 can be understood from realizing that equation (2.4) can be written as

$$p_X(x) = C\left(\frac{x}{x_0}\right)^{-\tau(x)-1}, \tag{2.9}$$

where $x_0 = \mathrm{e}^a$, $\tau(x) = \ln(x/x_0)/(2\sigma^2)$, and $C = \left(2\pi\sigma^2 x_0^2\right)^{-1/2}$ is a normalization constant (Montroll and Shlesinger, 1982).

The exponent $\tau(X)$ is a random variable normally distributed $\mathcal{N}(0, 1/(4\sigma^2))$. As σ increases, $\tau(X)$ tends to zero, and a lognormal resembles more and more a Pareto distribution like equation (2.2), with exponent $\tau \to 0$. Thus, we should not be surprised to observe Lévy-type behavior for finite values of N. This behavior can be regarded as a *truncated* Lévy Flight (Mantegna and Stanley, 1994), which means that even though equation (2.4) has all its moments finite, a finite sum of random variables resembles the Lévy Case presented in the previous section for small value of N.

Romeo *et al.* (2003) have shown that when $p_X(x)$ is a heavy-tail lognormal distribution $\mathcal{LN}(a, \sigma^2)$ with $\sigma \geq 1$, the convergence of the Law of Large Numbers can indeed be very slow.[6] They show that the typical (i.e., most probable) value Y_N^t of the sum in equation (2.5) can scale more than proportionately with the number of summands, for some *finite* range of N. That is, $Y_N^t \propto N^\beta$, with $\beta > 1$.

[6]Still, since all the moments of a lognormal distribution are finite, the Law of Large Numbers dictates that as N goes to infinity then $Y_N/N \to \mu = \mathrm{e}^{a+\sigma^2/2}$. That is, the typical value will eventually scale proportionately with N.

2.5.2 Constraints on the Population Size

The mathematical results we use here to understand the interplay between the moments of a distribution and the sample size required to assess the convergence to the LLN are the bounds on probabilities that the moments impose.

Recall that we want equation (2.7) to tell us how large N should be depending the characteristics of the distribution of $X_1,\dots,X_N$. If $\mathrm{E}\left[X_i^r\right]<\infty$ for some $r\geq 2$ we can combine Markov's and Rosenthal's inequalities to relate equation (2.7) with $\mathrm{E}\left[X_i^r\right]$ (see appendix E for the details of this derivation). This yields

$$\left[\frac{\alpha\varepsilon^r}{c(r)}\right]N^{r-1}\geq \mathrm{E}|X_1-\mu|^r+\left(\frac{\mathrm{Var}[X_1]}{N}\right)^{r/2}, \tag{2.10}$$

for some constant $c(r)$ (see Ibragimov and Sharakhmetov, 2001, 2002 for the values of this constant).

Equation (2.10) establishes an interplay between N and the moments of the individual random variables X_i in the population, *given* a confidence $1-\alpha$ that our per capita measure Y_N/N is within an ε from the population mean μ. This inequality, however, is not the sharpest possible. The reason is because Markov's inequality is not a sharp inequality. In spite of this, it does give us an indication of how N must scale. In this condition r is a free parameter. The best bound is for some value of $r^*\geq 2$:

$$N^*_{\mathrm{min}}\sim\left(\frac{c(r^*)}{\alpha\varepsilon^{r^*}}\right)^{\frac{1}{r^*-1}}\left[\mathrm{E}|X_1-\mu|^{r^*}+\left(\frac{\mathrm{Var}[X_1]}{N^*_{\mathrm{min}}}\right)^{r^*/2}\right]^{\frac{1}{r^*-1}}. \tag{2.11}$$

Let us apply equation (2.10) to our null model. First, we must recall that the moments of a lognormal distribution $\mathscr{LN}(a,\sigma^2)$ are given as

$$\mathrm{E}\left[X^k\right]=\mathrm{e}^{ka+k^2\sigma^2/2}. \tag{2.12}$$

Thus, in a lognormal distribution the moments increase exponentially with the square of the moment order k. Given this, the best bound in equation (2.10) will be for $r^*=2$ (according to Ibragimov and Sharakhmetov, 2002, $c(2) = 1$). Taking $\alpha = 0.1$, $\varepsilon = 0.1$, $\mu = 1$, $\sigma = 4.5$, we get that

$$0.001\, N \geq 6.23 \times 10^8 \left(1 + \frac{1}{N}\right). \tag{2.13}$$

We find that $N^*_{\min} \approx 6.23 \times 10^{11}$ is the minimum value that fulfills the condition. This is larger by two orders of magnitude than the value we found according to the numerical simulations in fig. 2.6 because of the aforementioned reasons. Yet, it highlights once more the effect of large heterogeneities in the validity of per capita measures.

2.6 Summary and Discussion

The motivation for this chapter started with the question about the causes of the wide differentials in per capita productivity across urban areas in the U.S., as shown in fig. 2.1. In particular, two stylized facts stood out: (i) a slight, yet statistically significant, increase in the average per capita productivity with population size, and (ii) a wide variance around this average. Two different, but not incompatible, interpretations about the root causes of these two observations were mentioned.

In the conventional interpretation, some factors of production, yet to be discovered, would account for the wide differences in productivity per capita. The slight effect of population size could be either due to an agglomeration effect, or it could also be that population size stands as a confounding variable of another effect that is stronger in larger cities than in smaller ones, as argued by Shalizi (2011).

In contrast, we provided an alternative interpretation in which the wide fluctuations could be due to an underlying stochasticity in the process of production in cities. Features (i) and (ii) can arise, for example, from the interplay between a broad heterogeneity in the productivity of individuals and the finiteness of cities. Following this latter interpretation, we introduced a null model of an urban system based on the assumption that individual productivity was independently drawn from the same lognormal distribution for all individuals across all cities. This null model helped us prove the point that a simple stochastic model of a city can reproduce the qualitative features of fig. 2.1.

The null model showed that when the Law of Large Numbers is violated the cross section of a finite realization of cities displays a superlinear relation between total productivity and population size. In the notation of linear regression models, if $y_1, y_2, \ldots, y_m$ are m realizations of the random variable Y_N for different N taken from $\{n_1, n_2, \ldots, n_m\}$, then

$$\widehat{y_i} \propto n_i^{\beta}, \quad \text{for } i = \{1, \ldots, m\}, \tag{2.14}$$

where the hat symbol $\widehat{\ }$ means the estimated value. Equation (2.14) is only observed when the city sizes are small relative to a number that defines the region of convergence of the LLN. In other words, it holds when $\max\{n_1, n_2, \ldots, n_m\} < N^*_{\text{min}}$, for some value N^*_{min} determined by the distribution describing the productivity of individuals, given by equation (2.11).

This all means that, methodologically, when there are broad inequalities and sizes are small (in the order of millions, perhaps) per capita transformations can give misleading information about the average individual productivity. Larger places can be more productive in per capita terms than smaller places, even when intrinsically they are not. And this means that, in our null model, population size is not just a parameter to control for size. Rather, population size is the variable that determines the superlinear growth of the aggregate productivity, not of the individual productivities. If in complex systems "the whole is more than the sum of the parts", in our null model, "the whole is typically more than the sum of the expected mean of the parts".

We did not derive the null model from economic principles. There are, however, several models to generate fat-tailed distributions of productivity. In appendix D, for example, we describe a simple stochastic model that can generate a variety of fat-tailed distributions of productivity based on the so-called Kesten process. Banerjee and Yakovenko (2010) have developed a model that generates a distribution of income that is exponential in the body and decays as a power-law in the tail. The assumption that the distribution of human productivity is lognormal, moreover, has empirical support (see Kleiber and Kotz, 2003, pp. 126–130). In the context of scientific output, one of the first to recognize a lognormal distribution describing the productivity of individuals was Shockley (1957). Shockley also provided some hypothesized mechanisms to explain the lognormality of productivity. Some years before, similar arguments about productivity were given by Roy (1950), although more intuitive and from an economic point of view. Shockley, Roy, and others (Aitchison and Brown, 1957), all invoke a multiplicative stochastic process determining socio-economic quantities. Hence, in the process of production, several substasks need to be achieved *in conjunction*, and depend on workers having different skills that also act multiplicatively. As a consequence, productivity results in a lognormal distribution as a consequence of applying the central limit theorem to the logarithms of the random variables representing each substep in the process (see Kleiber and Kotz, 2003 for other distributions). The lognormal distribution is going to prove important in subsequent chapters, in which suggestions that production at the level of the whole city will emerge.

A strong assumption of the null model is that the production process occurs solely at the level of individuals and independently of one another. Even though one can argue that the broad aspect of the lognormal distribution of productivities is, in effect, a result of interaction effects between individuals, this argument goes against the assumption of independence. If fact, cities are essentially characterized by their tight web of interconnected elements (Bettencourt, 2013). We should expect this interaction to correlate the productivity of individuals. As a result, it is more reasonable to think that the production process occurs at higher levels of organization, such as small groups of people, universities, firms, industrial conglomerates, and even cities as a whole. In these conditions, one could still observe $\widehat{Y_N} \propto N^{\beta}$ at the level of cities without a violation of the law of large numbers. Yet, the null model presented here is not entirely incompatible with this.

In cities, where interaction effects are strong and broad heterogeneities are the norm, scaling relationships of total productivity with population size can thus arise from two separate effects: one in which most individuals produce more on average by virtue of interacting in groups (e.g., agglomeration economies), and another in which a single individual produces a significant share of the total production of a city. We could label them as the "Jacobs" and "Edison" effects, respectively, given what Jane Jacobs wrote about the benefits of diverse human agglomerations, and given the productivity of Thomas A. Edison as a prolific inventor who must have raised significantly the per capita productivity of the former city of Menlo Park in New Jersey[7], with his laboratories.

Patents is precisely an interesting case in which both effects might be present and, more importantly, measurable. Recall that in fig. 2.3 we were able to reproduce the scaling exponent of patent applications with population size in U.S. Metropolitan and Micropolitan Statistical Areas in the year 2000. We tuned the parameter σ of the lognormal distribution to $\sigma = 6.0$ and we generated the scaling exponent $\beta = 1.37$ that is empirically observed. But there are other possibilities of explaining fig. 2.3. Namely, we can decompose the exponent as:

$$\widehat{y_i} \propto n_i^{1+\delta+\eta}. \tag{2.15}$$

The exponent δ could come from a Jacobs effect, which urban scaling theory predicts $\delta = 1/6$ (Bettencourt, 2013), and the remaining $\eta \approx 0.20$ could be explained, possibly, by

[7]The name was changed to Edison in 1954 in honor to the inventor.

an Edison effect from a lognormal distribution with $\sigma \approx 4.7$ (see fig. 2.5).[8] Future work is needed to disentangle these effects.

Our null model also generates large fluctuations around equation (2.14), which qualitatively match the fluctuations observed in fig. 2.1 and fig. 2.2. This result is not surprising since the null model is set to work outside the range of validity of the LLN, and thus large fluctuations are expected. But for these same reasons, it is a non-trivial problem to analytically describe these fluctuations. The question here is what is the probability distribution of the random variable $Y_N = \sum_{i=1}^{N} X_i$, for a fixed N, when $X_i \sim \mathscr{LN}(a, \sigma^2)$ are i.i.d. random variables. It is an open problem with important applications in other fields such as engineering, and it has received much attention in the literature (Szyszkowicz and Yanikomeroglu, 2009; Mehta *et al.*, 2007; Beaulieu *et al.*, 1995). Yet, no closed-form is known for even the simplest case of $N = 2$.[9]

2.7 Concluding Remarks

In this chapter, we provided a set of computational and mathematical results concerning the statistics of productivity measures. In particular, we showed that the assumptions be-

[8] Bettencourt *et al.* (2007b) show that the productivity per inventor does not increase with population size. This does not invalidate the hypothesis in equation (2.15) since we can view the production process occurring at a higher organizational level. In that case N stands not as the number of people, but as the number of higher-level groups of production in a city, such as firms and universities, and the random variable X would stand as the number of inventors in each group. In this way, the aggregate number of inventors per city Y_N would grow superlinearly with population size, if the number of firms and universities are proportional to the number of people.

[9] We know, however, the two boundary conditions:

$$\frac{Y_N}{N} \sim \begin{cases} \mathscr{LN}\left(a, \sigma^2\right) & \text{for } N = 1, \\ \mathscr{N}\left(\mu, (e^{\sigma^2} - 1)\mu^2\right) & \text{for } N \to \infty, \end{cases} \tag{2.16}$$

where $\mu = e^{a+\sigma^2/2}$.

hind uses of per capita measures might not hold and that aggregate measures of output are preferred, depending on the interplay between city sizes and the distribution of individual productivities. Moreover, we showed that this interplay determines the overall statistical properties of the aggregate output. Hence, studies of the statistical properties of aggregate output must be carried out if our aim is to understand the mechanisms generating wealth.

Our present study also highlights the importance of understanding the statistical properties that describe individuals within cities. Social systems, where interaction of a finite number of elements is the defining factor, have been conspicuous in displaying broad and heavy-tailed distributions, as opposed to narrow and thin-tailed distributions. This heterogeneity and the inequalities that it translates to when speaking of productivity, wages, income, and wealth has consequences for the welfare of society. But as we have shown here, it can determine the overall profile of productivity of a whole urban system of cities.

Note: the recent article by Sornette et al. *(2014) examined some ideas similar the ones mentioned in this chapter.*

Chapter 3

THE STATISTICS OF URBAN SCALING AND THEIR CONNECTION TO ZIPF'S LAW

As was shown in the previous chapter, distributional considerations regarding aggregate urban metrics are important both for the understanding of what occurs within cities as for the understanding of the urban system as a whole. However, the statistics of urban indicators have not been analyzed in detail in previous studies, raising important questions about the full characterization of urban properties and how scaling relations may emerge in these larger contexts. Here, we build a self-consistent statistical framework that characterizes the joint probability distributions of urban indicators and city population sizes across an urban system. To develop this framework empirically we use one of the most granular and stochastic urban indicators available, specifically measuring homicides in cities of Brazil, Colombia and Mexico, three nations with high and fast changing rates of violent crime. We use these data to derive the conditional probability of the number of homicides per year given the population size of a city. To do this we use Bayes' rule together with the estimated conditional probability of city size given their number of homicides and the distribution of total homicides. We then show that scaling laws emerge as expectation values of these conditional statistics. Knowledge of these distributions implies, in turn, a relationship between scaling and population size distribution exponents that can be used to predict Zipf's exponent from urban indicator statistics. Our results also suggest how a general statistical theory of urban indicators may be constructed from the stochastic *dynamics of social interaction processes in cities.*[1]

3.1 Introduction

The search for a general multidisciplinary science of cities is a fundamental scientific problem with strong roots in economics (Fujita *et al.*, 2001; O'Sullivan, 2006), sociology (Fischer, 1984; Flanagan, 2010; Mumford, 1961), urban planning and architecture (Bacon, 1976; Hall, 1976). As human populations become increasingly urban the quantification of general insights and solutions that transcend each particular place is increasingly necessary and would have important consequences for our fundamental understanding of human societies and for urban planning and policy (Bettencourt and West, 2010).

Cities should be regarded primarily as dynamical social networks, constantly changing in terms of their composition and interactions. Consequently, urban indicators, denoted by Y, and population N, should be treated in general as stochastic variables. More specifically, there are practical circumstances when a full statistical approach to urban quantities becomes necessary. For example, a statistical treatment of urban indicators is essential when average characterizations is insufficient because of the granularity that arises when dealing with small integer numbers in Y (or in N). In such extreme regimes we may investigate if and to what extent urban scaling laws apply and how they may emerge in the limit of large numbers, when Y can be thought of as an effectively continuous variable.

In order to probe urban indicators that show granularity and a large level of temporal and geographic variation we analyze here data on annual homicides in cities of three Latin

[1]This chapter was published (Gomez-Lievano *et al.*, 2012) in collaboration with Dr. Hyejin Youn and Professor Luís M.A. Bettencourt of the Santa Fe Institute. We acknowledge José Lobo and Geoffrey West for discussions and for comments on the manuscript. We thank Diego Valle for data on homicides for Mexican municipalities and for useful discussions. We also acknowledge Jesse Taylor for helpful comments and suggestions, and Maria Jose Uribe for discussions and for providing us help with the Colombian data.

American countries over a several year period during which national homicide rates have varied substantially. We analyze data from three of the largest nations in Latin America, presently showing some of the highest homicide rates in the world: Brazil, Colombia and Mexico, for which data are available at the municipal level.

The number of homicides is a quantity that is widely available at the local level in developed and developing nations. It is thought generally to be reliably reported, notwithstanding some important caveats (Valle-Jones, 2011). For these reasons, we use the annual number of homicides in Latin American cities to develop a statistical approach to urban scaling.

Homicides, as the ultimate expression of violence in human societies, are a widely investigated quantity (The World Bank, 2011a,b). Many reasons have been advocated for the rise and fall of homicides in cities throughout the world, especially in the US and Europe (Pinker, 2011). Here it is not our intention to distinguish between these ideas or propose new ones, but to determine general characteristics of the statistics of homicides in connection with the population size of a city. More specifically, our main objective is to establish general properties of the statistics of urban indicators in the limit of high granularity and to investigate if and how urban scaling laws emerge and are related to Zipf's law for the population size of cities. We expect that such insights should extend to other urban indicators and shed some light on a full statistical theory of cities in terms of their quantitative observable properties.

In the next section we discuss some of the characteristics of the data and our main formal objective, the estimation of the conditional probability density $P_{Y|N}(y|n)$ for a particular realization Y in a city with population N. Because no two cities have the same population direct estimation of this probability is impractical so we exploit Bayes' rule to compute instead $P_{N|Y}(n|y)$ and $P_Y(y)$. We describe the statistical properties of these two distributions and then derive a closed form for $P_{Y|N}(y|n)$. We then show that scaling laws emerge as the expectation value of Y given N and how knowledge of the conditional distributions and of $P_Y(y)$ lead to Zipf's law for the size distribution of cities. Finally, we discuss several qualifications and generalizations of these results and some of their general implications.

3.2 Results

3.2.1 Scaling Relations and Units of Analysis

Bettencourt *et al.* (2007a, 2010) have recently shown that many urban properties Y_i vary, on average, with city population size N_i according to a scaling relation

$$Y_i(t) = Y_0(t)\, N_i(t)^{\beta}, \tag{3.1}$$

where the subscript i refers to a particular city within an urban system at time t, $Y_0(t)$ sets the baseline value of Y for the urban system and the exponent β measures the average relative change in Y with N, $\beta = (\Delta \ln Y)/(\Delta \ln N)$. In particular, for socioeconomic quantities such as urban GDP, wages or violent crime β is typically superlinear ($\beta > 1$), expressing an average per capita increase in these quantities with city size N. Here we go beyond mean expectations to show how equation (3.1) emerges statistically.

We have also observed that for US metropolitan areas many urban quantities vary only slowly, with most change being due to the temporal variation of Y_0 and to the dynamics of population change. This has the consequence that deviations from average scaling—for example in economic quantities or measures of innovation—tend to persist, and sometimes be reinforced, for several decades (Bettencourt *et al.*, 2010). Under these circumstances it becomes difficult to observe systematic variations in the statistics of urban metrics, precluding us from eventually establishing the properties of their underlying processes.

To address these points, we introduce here new extensive data sets for homicides in three fast evolving (and developing) nations: Brazil, Colombia and Mexico. These nations are presently among the most violent in the world with registered homicide rates greater than 15 per 100,000 inhabitants, see The World Bank (2011a,b). Their homicide rates have, in addition, experienced substantial changes over time, both at the national level and in some particular cities. In all three cases the rise in violence, especially in particular cities, has become a major impediment to national economic development and a challenge to international security. Changes in crime rates in these nations, as elsewhere, have been attributed to new initiatives to fight organized crime (Goertzel and Kahn, 2009) or to the rise of several organized crime groups and to 'wars' between them (Gaviria, 2000). Although these and other explanations for the variation of crime in cities have been advanced and are widely discussed in the literature, the evaluation of their relative merits requires, in our opinion, improved statistical models, that quantify and specify the nature of fluctuations and go beyond average rates.

In Brazil, Colombia and Mexico the smallest spatial unit for which data are available is the municipality (*municipio*). Municipios are defined as the smallest administrative units with a local government. Because municipalities partition the entire national territory, their interpretation as urban units is flawed, just as it would be to assume that each county in the US, for example, is a city. Most municipalities consist, in fact, of several human settlements over extensive rural areas. This introduces a limitation in the resolution at the smallest population scales. At the larger population scales we can address this issue because large functional cities (metropolitan areas) are made up of a set of municipalities. Thus, bearing in mind these caveats, we adopt a definition of urban units in terms of metropolitan areas for which an official definition exists, plus the remaining non-metropolitan municipalities. Data sources, definitions and more details are provided in the Methods section.

We motivate the need for our statistical study by displaying in fig. 3.1 the scaling of total homicides versus population size over a single year. The solid line fits the scaling of homicides for metropolitan areas only. Large differences are displayed between municipalities, and our goal is to characterize these fluctuations in a complete framework. We will not discuss the specificities of urban homicides but their general statistical nature, and their relation to scaling and Zipf's law.

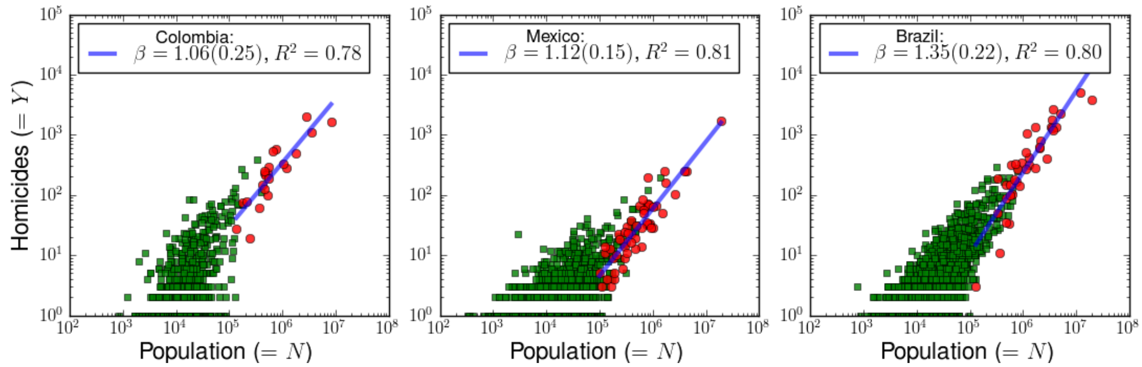


Figure 3.1: Annual Number of Homicides in Cities of Colombia, Mexico and Brazil Versus Population Size (2007). Large Cities Are Defined in Terms of Metropolitan Areas which Are Aggregations of Municipalities (Red Circles) While Non-Metropolitan Municipalities Are Shown Separately (Green Squares). The Solid Blue Line Fits only the Scaling of Homicides for Metropolitan Areas. Large Variations, Especially Among the Smaller Population Units, and the Fact that Many Municipalities Have $Y = 0$ (Not Shown) Prevent a Direct Scaling Analysis. However, it is Possible to Analyze the Data Consistently Through the Estimation of Conditional Probabilities.

3.2.2 Bayesian Approach to the Statistics of Urban Indicators

Equation (3.1) is an average statement that cannot be obeyed exactly in every instance. This is not only because all cities have specific local characteristics and urban indicators fluctuate over time but, more fundamentally, because a continuous scaling relation must break down in the limit of small discrete numbers. The correct statement must then be formulated in probabilistic terms. To do this we think of both Y and N as stochastic variables, and of their values at each particular city and time as statistical realizations. We can then estimate their probability distributions.

This problem is specified in terms of the conditional probability distribution function of $Y = y$, given a city of population $N = n$, $P_{Y|N}(y|n)$. We use Bayes' rule

$$P_{Y|N}(y|n) = \frac{P_{N|Y}(n|y)\ P_Y(y)}{P_N(n)}, \tag{3.2}$$

to compute it, given knowledge of the probability distribution $P_Y(y)$ of homicides in cities regardless of their population, and the conditional probability distribution $P_{N|Y}(n|y)$ for the population size of cities with a given number of homicides. The denominator is a constant in y and can be expressed as the trace of the numerator over all values of Y. We will return to this point below as $P_N(n)$ is Zipf's probability density function for city population sizes.

The reason to estimate $P_{Y|N}(y|n)$ indirectly is motivated primarily by practical considerations. To estimate $P_{Y|N}(y|n)$ directly we would have to aggregate cities of similar size together into arbitrary discrete size intervals (binning), potentially introducing errors and leading to several additional complications. To avoid this, we exploit to our advantage the granularity of the data as there are substantial numbers of cities with $Y = 0, 1, 2, 3, \ldots$ leading naturally to estimates of $P_{N|Y}(n|y)$.

3.2.3 Estimating the Distribution of Total Urban Homicides

The distribution of total homicides in cities $P_Y(y)$ must reflect the fact that urban properties change (super)extensively with population and that there are cities with widely varying sizes. As such we should expect $P_Y(y)$ to be a broad distribution. Because of these general facts, power-law probability densities (Zipf or Pareto distributions), are common among urban metrics. More specifically, these distributions account for the fact that a small number of cities are responsible for most homicides and that a large number of cities display only a few. In Mexico, for example, approximately 60% of homicides come from 2% of cities! Similar numbers characterize Colombia and Brazil for the years studied.

In practice, we adopt a common procedure of plotting the complementary cumulative distribution function rather than the probability density function, which avoids the noisy character of the tail for large cities.

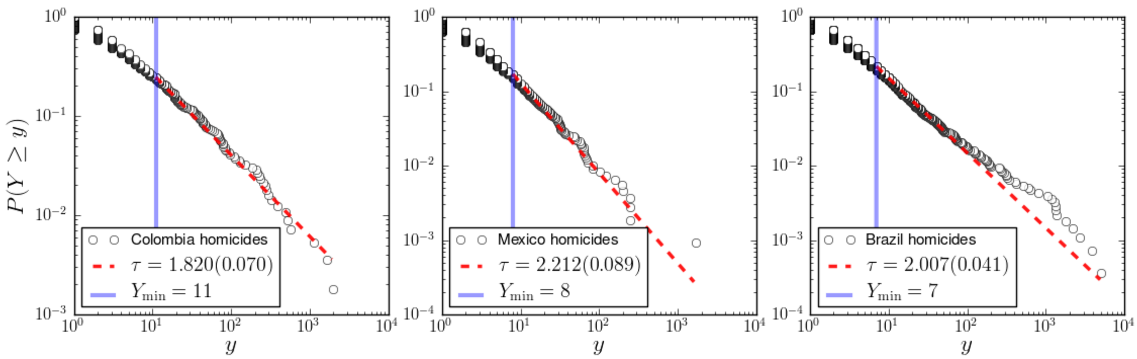


Figure 3.2: Cumulative Normalized Distributions of Homicides in Colombia, Mexico and Brazil (2007) Are Well Described by Power-Law Distributions. Here We Plot Not the Density Function but the Complementary Cumulative Distribution to Attenuate the Tail Fluctuations and Ease Visual Interpretation. Best Fits (Dashed Red Line) of the Form $P(Y \geq y) = Cy^{-\tau+1}$ Were Estimated Using the Procedure in Clauset *et al.* (2009) to the Density Function (See Methods Section). Standard Errors Are Reported in Parenthesis. The Solid Blue Line Shows the Minimum Value of Y for which a Power-Law Fit Holds. While the Distribution of Total Homicides is Scale Invariant, this is the Result of Tracing More Predictable Conditional Distributions for Each City Over a Broad Distribution of City Sizes (See Text).

The empirical cumulative distributions of homicides for the year 2007 in Colombia, Mexico and Brazil are shown in fig. 3.2. These distributions appear very similar, showing a heavy tail for several decades and an effective lower cutoff for small values of Y. We were unable to reject power-law fits using the procedure of Clauset *et al.* (2009). We assumed the functional form of $P_Y(y)$ to be

$$P_Y(y) = \frac{(y+k)^{-\tau}}{\varsigma(\tau,k)}, \quad y \in \mathbb{N}, \tag{3.3}$$

where $\tau > 0$ is the power-law exponent and k is a positive real number, which allows $P_Y(y)$ to remain analytic as $y \to 0$. Here

$$\varsigma(\tau,k) = \sum_{y=0}^{\infty} (y+k)^{-\tau} \tag{3.4}$$

is the generalized or Hurwitz zeta function Olver *et al.* (2010), which ensures the normalization of $P_Y(y)$ as a discrete variable.

3.2.4 Estimating $P_{N|Y}(n|y)$ and Deriving $P_{Y|N}(y|n)$

To calculate $P_{N|Y}(n|y)$ we fix the value of $Y = y$ and estimate the probability distribution over population. Figure 3.3 shows the histograms of frequencies of homicides for Colombia, Mexico and Brazil, for a range of Y. Note that N in the x-axis is plotted on a logarithmic scale ($\ln(N)$). These figures give us an impression of what type of probability distribution describes the data. We observe that all distributions, at each value of Y, show a distinct peak with definite mean and variance. The null hypothesis of a Poisson distribution was rejected with high confidence by a maximum likelihood method (see appendix H). Instead, these are well fit in terms of a lognormal distribution:

$$P_{N|Y}(n|y) = \frac{1}{n\sqrt{2\pi\sigma_y^2}}\, \mathrm{e}^{-\frac{(\ln n - \mu_y)^2}{2\sigma_y^2}}, \tag{3.5}$$

where the subscript in μ_y and σ_y indicates that these parameters are in general functions of y.

The shape of this distribution, which Bettencourt *et al.* (2010) had noted more implicitly for other quantities, is perhaps curious, first, because it does not conform to the more classic distributions, such as the Gaussian or Poisson, despite the fact we are dealing with count data that are traditionally related to neutral processes like the law of rare events (see Frank, 2009). And second, because it states that urban metrics are much more predictable given other variables (here simply population size) than a Zipfian distribution might have

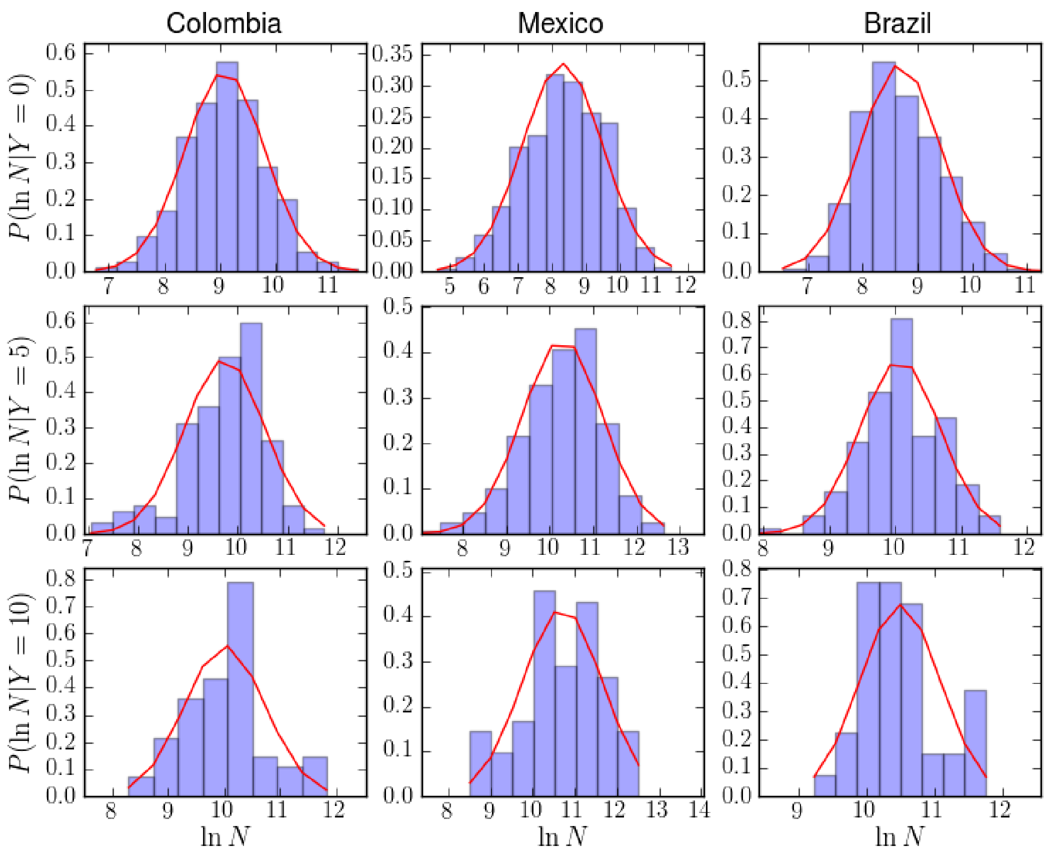


Figure 3.3: Normalized Frequency Histograms of the Logarithm of City Population for Varying Number of Observed Homicides Y. Each Column Corresponds to a Different Country and Each Row, From Top to Bottom, Corresponds to the Values $y = 0, 5$ and 10 Homicides Per Year. A Lognormal Distribution (Notice the x-axis Is Expressed in Terms of $\ln N$) Is Shown as a Solid Red Line, with Parameters Obtained Via Maximum Likelihood Estimation.

lead us to believe. Thus, effectively a Zipf distribution blurs fairly predictable quantities, given N, over a broad range of population sizes. Seen from the opposite perspective, lognormal distributions are what we observe if we look at the variables described by a Zipf distribution through a “lens” that allows us to distinguish its many (and widely varying in size) component units (cities).

One drawback of the lognormal distribution is that both N and Y are in reality discrete random variables, whereas the lognormal describes typically a continuous stochastic variable. (Discrete lognormal distributions are sometimes used in the statistics literature, see Anscombe, 1950 and references therein). In spite of this property, it is still reasonable to assume that the variation in population is approximately continuous as the minimal values of N are typically on the order of thousands.

The mean and variance are given by:

$$\langle N \rangle_y = \mathrm{e}^{\mu_y + \sigma_y^2/2} \tag{3.6}$$

$$(\Delta N_y)^2 = (\mathrm{e}^{\sigma_y^2} - 1)\mathrm{e}^{2\mu_y + \sigma_y^2}. \tag{3.7}$$

The maximum likelihood estimators of the lognormal parameters are:

$$\widehat{\mu_y} = \frac{1}{n_y} \sum_{i \in S_y} \ln N_i \tag{3.8}$$

$$\widehat{\sigma_y^2} = \frac{1}{n_y} \sum_{i \in S_y} (\ln N_i - \widehat{\mu_y})^2, \tag{3.9}$$

where n_y is the number of cities in the set S_y with y homicides.

If the normal distribution holds in terms the logarithmic variables of population given different values of Y, we can collapse the different histograms of fig. 3.3 by standardizing log-variables. We achieve this by calculating the maximum likelihood estimators of the mean and variance for every value of Y, and then plotting in the same histogram the distribution for several values of Y. Figure 3.4 shows these standardized distributions. This procedure has its limitations due to the fact that as we increase Y, the number of cities decreases, until there is only one city with given Y and N and statistical estimation becomes impossible. Conversely, it has the advantage that the shape of the distribution $P_{N|Y}(n|y)$ for several values of Y can be displayed in one single figure.

We can now estimate the parameters of equation (3.5) using section 3.2.4 and section 3.2.4, and plot $\widehat{\sigma_y^2}$ (see fig. 3.5) and $\widehat{\mu_y}$ (see fig. 3.6) versus Y, to infer their functional Y-dependence.

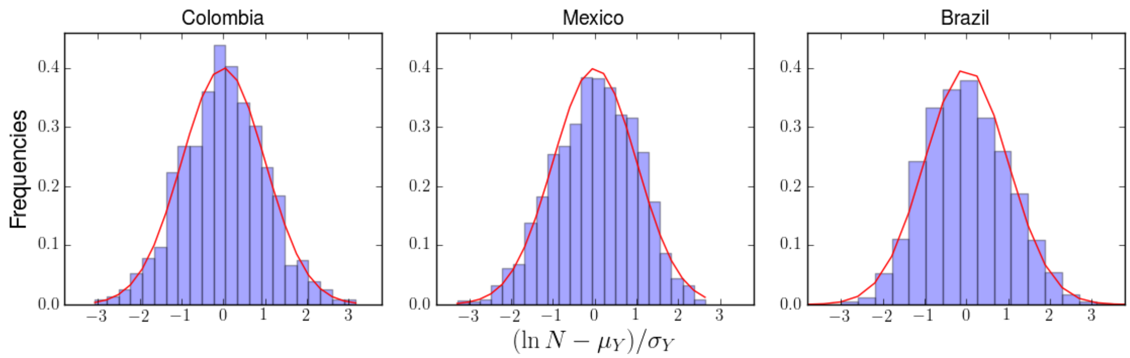


Figure 3.4: Collapsed Histograms of $P_{N|Y}(n|y)$ Across Values of Y in 2007. Lognormal Probability Density Functions for the Three Nations Are Shown as Solid Red Lines. This Shows that Power-Law Distributions Describing Total Homicides in the Urban Systems Have in Fact More Predictable Statistics When Conditioned on City Population Size.

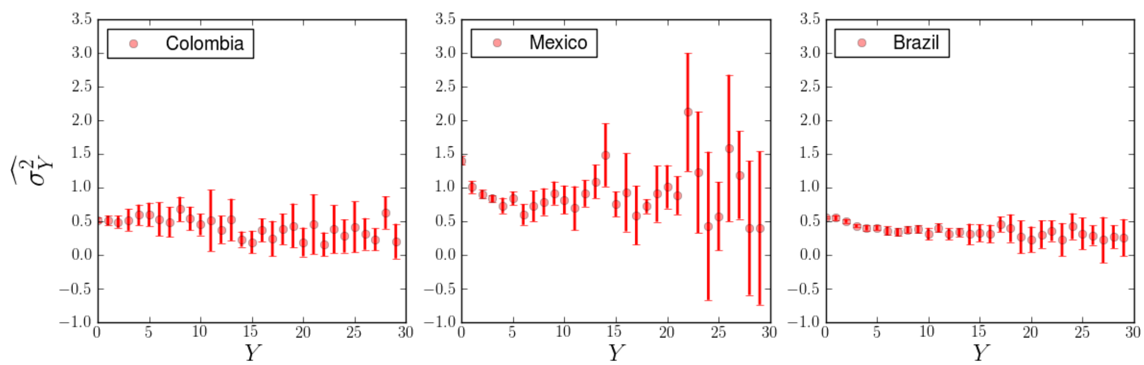


Figure 3.5: Estimates of σ_y^2 (Via Maximum Likelihood) for Different Values of $Y \in \{0, \ldots, 29\}$, for Colombia, Mexico, and Brazil. A Different Curve Was Constructed for Every Year of the Analysis (See Methods). The Plots Show the Average Over Several Years. Error Bars Represent One Standard Deviation Intervals (67% Confidence Level). The Plots Show no Clear Systematic Y-Dependence of $\widehat{\sigma_y^2}$. This Suggests, in Turn, that Each Country has a Characteristic Variance of its Indicators Conditioned on Other Urban Quantities. In this Respect, it is Interesting to Note the Similarities Between Colombia and Brazil.

The behavior of σ_y^2 shown in fig. 3.5 is stable and we will assume it to be constant henceforth. Because of this we can reject other count models such as the Negative Binomial, which is designed to model over-dispersed data. The curves shown in fig. 3.6 display a logarithmic growth of $\widehat{\mu_y}$ on y. The most general logarithmic function that can be fit to

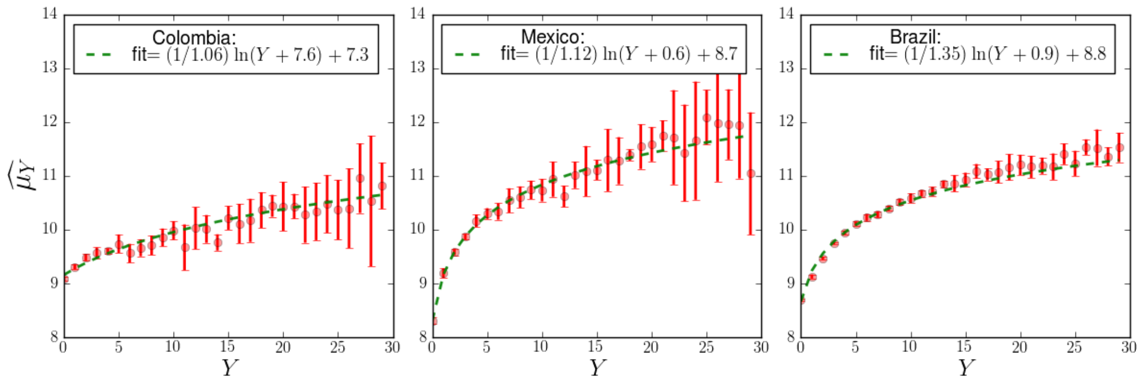


Figure 3.6: Estimates of μ_y (Via Maximum Likelihood) for Different Values of $Y \in \{0,\ldots,29\}$, for Colombia, Mexico, and Brazil. A Different Curve Was Constructed for Every Year of the Analysis, and the Points Plotted Are the Averages Over Several Years. The Error Bars Represent One Standard Deviation Intervals about the Mean. Plots Show a Logarithmic Dependence on Y, from which a Scaling Relationship Emerges in Terms of Expectation Values (See Text). Best Fits Were Obtained Using a Levenberg-Marquardt Algorithm, Weighting Every Point by its Error (See Methods).

$\widehat{\mu_y}$ (see fig. 3.6) is

$$\widehat{\mu_y} = f(y) = b\ln(y+r) + \ln A \tag{3.10}$$

where r is a positive constant that allows the logarithm to remain finite (and positive) as $Y \to 0$ and A is a positive number. Below, the constant b will be identified with the scaling exponent $1/\beta$. This is the reason why the values of these parameters in fig. 3.1 and fig. 3.6 coincide. The rest of the paper rests on these two assumptions about the behavior of σ_y^2 and μ_y, suggested by fig. 3.5 and fig. 3.6. For the fitting procedure of the remaining parameters see the Methods section.

Finally, using equation (3.2), we derive the conditional probability function $P_{Y|N}(y|n)$. If equation (3.5) holds for all $Y \geq 0$, using equation (3.3), we obtain

$$P_{Y|N}(y|n) \propto (1/\widetilde{y})\exp\left[-\frac{1}{2\sigma_y^2}(\ln n - \mu_y)^2 + (1-\tau)\ln\widetilde{y}\right], \tag{3.11}$$

where $\widetilde{Y} \equiv Y + k$.

Using equation (3.10) to replace μ_y for $\widehat{\mu_y}$, we obtain

$$P_{Y|N}(y|n) \propto (1/\widetilde{y}) \exp\left[-\frac{1}{2\sigma_y^2}\left(\left(\ln n - \ln A\,(y^*)^b\right)^2 + 2\sigma_y^2(\tau-1)\ln\widetilde{y}\right)\right]. \tag{3.12}$$

We can expand the squared terms, the logarithms, group some of the terms, so this equation transforms into:

$$P_{Y|N}(y|n) \propto (1/\widetilde{y}) \exp\left[-\frac{1}{2\sigma_o^2}\left(\ln^2 y^* - 2\,\ln\left(\frac{(y^*)^P}{\widetilde{y}^{\sigma_o^2(\tau-1)}}\right) + P^2\right)\right], \tag{3.13}$$

where $y^* = y + r$, $P = \frac{1}{b}\ln(n/A)$ and $\sigma_o = (\sigma_y/b)$.

Now, recall that both r in $y^* = y + r$ and k in $\widetilde{y} = y + k$ were introduced to account for the limit when $y \to 0$. These constants generate the expected limits and prevent us from dividing by zero in the power-law distribution and from taking the logarithm of zero in $\widehat{\mu_y}$. There are no constraints that keep us from assuming them to be equal (and from considering them to be small). Indeed, both introduce a characteristic scale which manifests itself as a regime change in the scaling behavior when cities are very small and realizations of zero homicides (or other discrete measures) begin to occur. Therefore, it is not unreasonable to assume they are they same, thus $Y^* \approx \widetilde{Y}$ (see the Methods section for an estimation of k). Under this assumption we can complete the square and compute the posterior distribution. Realizing that $P_{Y|N}(y|n) = P(Y^*|N)$ because $\Delta Y^*/\Delta Y = 1$, and keeping only Y dependent terms (the others will ultimately be absorbed by the normalization constant), we arrive at

$$P_{Y^*|N}(y^*|n) \propto (1/y^*) \exp\left[-\frac{1}{2\sigma_o^2}\left(\ln y^* - (P - \sigma_o^2(\tau-1))\right)^2\right], \tag{3.14}$$

which is a lognormal distribution for Y^* given $N = n$, with parameters $\mu_n = P - \sigma_o^2(\tau - 1)$ and $\sigma_n = \sigma_o$. By expressing the distribution parameters in the original variables, and by introducing the proper normalization constant, we finally obtain

$$P_{Y^*|N}(y^*|n) = \frac{1}{y^*\sqrt{2\pi\sigma_n^2}}\,\mathrm{e}^{-\frac{(\ln y^* - \mu_n)^2}{2\sigma_n^2}}, \tag{3.15}$$

$$\mu_N = \frac{1}{b}\ln\left(\frac{n}{A}\right) - \sigma_n^2(\tau-1) \tag{3.16}$$

$$\sigma_n^2 = \sigma_y^2/b^2. \tag{3.17}$$

These expressions connect the lognormal statistics of the conditional distribution $P_{Y|N}(y|n)$ with scaling and Zipf's law for the size distribution of cities. As we show below this leads to a relationship between scaling and Zipf's exponents.

Determining these conditional distributions enables us to calculate their moments, such as the mean and variance. We take equation (3.10) and section 3.2.4 to derive $\langle N \rangle_y$ and $\langle Y \rangle_n$ explicitly in terms of y and n, that is:

$$\langle Y + r \rangle_n = \left(\frac{\mathrm{e}^{(3/2-\tau)\sigma_n^2}}{A^\beta} \right) n^\beta \tag{3.18}$$

$$\langle N \rangle_y = A\, \mathrm{e}^{\sigma_y^2/2}\, (y+r)^{1/\beta} \tag{3.19}$$

where $\beta = 1/b$ is recovered as the exponent of the scaling relation in equation (3.1). These two expressions represent complementary scaling relations. Note that they are not identical statistically as they express the expectation value of each variable in terms of a given value of the other, not its mean.

Similarly, the standard deviations ΔY_n^* and ΔN_y can be expressed as

$$\Delta Y_n^* = \Sigma_n\, n^\beta \tag{3.20}$$

$$\Delta N_y = \Sigma_y\, (y+r)^{1/\beta}, \tag{3.21}$$

where Σ_n and Σ_y are proportionality coefficients.

From the preceding sections it should already begin to be clear how the lognormal distribution relates to Zipf's law. We can show how a power-law distribution emerges by deriving the probability distribution of N. In equation (3.2), $P_N(n)$ is called the "evidence", and acts in practice as a normalization constant. It can be calculated from knowledge of the numerator as

$$\begin{aligned} P_N(n) &= \sum_{y^*=r}^{\infty} P(n|y^*)P(y^*) \\ &\propto n^{-\alpha}, \end{aligned} \tag{3.22}$$

which is a power-law distribution. It then follows that the various exponents are constrained to obey the relationship (see Methods section)

$$\beta = \frac{\alpha - 1}{\tau - 1}. \tag{3.23}$$

Thus, from this perspective Zipf's law follows from $P_Y(y)$ and the statistics of Y for cities of a given size and the observation of scaling of its expectation value.

If superlinear scaling ($\beta > 1$) holds for some urban indicator Y, we can predict population sizes to be power-law distributed with exponent $\alpha > \tau$, and vice versa if the scaling is sublinear. If $\alpha \approx 2$ (unity in a rank-size plot) as has been observed for several urban systems (Soo, 2005), superlinear scaling means that $\tau < 2$, and thus the quantity Y may lack a definite mean and variance. In these cases, references to "average cities" have no sound mathematical meaning. Note however that it is also possible that $\alpha > \tau > 2$ provided that Zipf's exponent is sufficiently larger than 2, as has been argued long ago by Mandelbrot (Mandelbrot, 1961) in a different context. In general these properties can be used to constrain the value of Zipf's exponent from the observation of the statistics of many different urban indicators and knowledge of their average scaling properties.

3.3 Discussion

In this chapter we characterized the statistics of homicides — a highly variable and granular metric — in three fast changing urban systems in Latin America. This analysis allowed us to address the statistics of scaling laws under extreme conditions and investigate how they emerge for noisy and granular variables within a larger probabilistic context.

We have found that homicides Y occurring in cities of Brazil, Colombia and Mexico all follow statistics that are well described by lognormal distributions. These distributions are parametrized by an expectation value that is population size dependent and a variance of the log-variables that is not (or that at least can be assumed not to be, for the data analyzed here). In this context scaling laws emerge as the expectation value of Y as a function of N, $\langle Y \rangle (N) \sim Y_0 N^{\beta}$.

The allometric scaling relationship, when expressed in terms of logarithms, exposes the issue that it cannot hold in the limit of Y or N going to zero (unless they do so together). We have devoted particular attention to this regime and found that effectively annual homicide rates saturate at a very small but non-zero value at sufficiently small N. In this sense true scale invariance emerges only when $Y \gg 0$. A dual scaling law for $\langle N \rangle (y)$ emerges from a Bayesian inversion of the relationship for Y and we have shown that this - i.e. the estimation of $P_{N|Y}(n|y)$ at small discrete Y - is often the most practical way to estimate $P_{Y|N}(y|n)$. This lead us in turn to the consideration and estimation of $P_Y(y)$ - the distribution of the total number of homicides across cities - which we found to have a Zipfian form. Because these distributions can be used to derive Zipf's law through marginalization, we obtained a relationship between urban indicator statistics, urban scaling laws and Zipf's distribution in the form of a constraint between the scaling exponent and the Zipfian exponents for Y and N.

Much effort has traditionally been devoted to model the broad distributions and lack of characteristic scales displayed by urban systems (Simon, 1955; Gabaix, 1999; Saichev *et al.*, 2010). However, our results show that parts of the urban system manifest greater predictability than is usually recognized. Although non-broad distributions would be expected to arise for many quantities when considering cities of fixed size, lognormal statistics are special because they point to multiplicative processes. If these processes depend on the structure of social interactions, lognormality then suggests that quantities should scale with city size in non-trivial ways.

Furthermore, the consistency between lognormal statistics for individual cities and Zipfian distributions for the urban system, as well as scaling relations across city sizes, suggest that local indicators are the result of self-consistent urban system dynamics and that these indicators are naturally bounded. Consequently, when considering goals for urban planning it is important to think at once locally and at the level of the urban system. In this light, on the one hand, questions about particular cities and the magnitude of their metrics may not make much sense unless we take into account the whole urban system in which they are embedded. On the other hand, characterizing urban systems only through power-law distributions prevents us from observing finer quantitative patterns present locally. Several mechanisms have been proposed for the emergence of lognormal (Sornette, 2006) and power-law (Frank, 2009; Baek *et al.*, 2011; Newman, 2005; Sornette, 2006) statistics, usually relying on multiplicative random processes (Montroll and Shlesinger, 1982; Redner, 1990). Population size dependent stochastic interaction processes within cities, which are multiplicative, provide a natural setting to explain these observations and will be the focus of future research.

Given the general implications of these results a few remaining issues and some caveats are worth further discussion. First of all, we motivated the lognormal distribution as a good general description of the data. However, the data may be compatible with other statistical densities, specifically a Laplace distribution (which is also characterized by two parameters a scaling mean and a fixed dispersion, see Bettencourt *et al.*, 2010). We found no consistent evidence in our empirical analysis that pointed conclusively to the need for these alternative and potentially more complex statistical models, but such need may arise as larger datasets are analyzed.

Second, one of our main results is the observation of deviations from scaling in the limit $Y \to 0$, where we are also dealing with small municipalities in terms of N. This regime and its statistical treatment is fraught with empirical difficulties, including the fact that we are then dealing predominantly with rural territories in which several small towns are aggregated together as a municipality. Thus, these units are not true single cities. To address this point more disaggregated data would be necessary to probe the behavior of Y

in truly small towns. In this sense our parametrization of the several distributions through the introduction of a saturating constant should be seen as provisional, and is in any case not unique. Another issue, that becomes important for small cities, is the use of annual homicides. If in reality the expected homicide rate vanishes only with vanishing population size, but becomes very small in small towns then it will take on average a longer and longer period of time for any homicides to be observed and any chosen time period will lead to an underestimation of such a rate for a suitably small city. Thus, by making the time period that defines the homicide rate longer we should see the saturating parameters decrease and scale invariance be restored to smaller and smaller population scales. We probed this regime empirically and indeed observed a systematic reduction in the size of cities with zero homicides, but the full consideration of this question is complex and is beyond the current analysis. The empirical and theoretical consideration and re-analysis of these issues may become possible in the future and would be interesting to pursue in order to investigate the limits of urban scaling in small population agglomerations. While it seems plausible to us that a finite probability of violence exists in human communities of any size, the lower limit may be difficult to probe in practice.

As they stand the present results suggest several interesting new questions for future research. First, they provide a mesoscopic view of urban indicators and take a step in suggesting the form of a statistical mechanics approach to universal aggregate properties of cities, such as scaling laws and size distributions. Such an approach should lead to theory and methods to bridge scales of analysis from individuals, through social and economic organizations, to entire cities and urban systems.

Finally, it is interesting to briefly discuss the practical implications of the statistical treatment of urban indicators developed here. Quantitative knowledge of the distribution of indicators for a given population size allows us to make predictions e.g. for the homicide rate of a particular place with quantified levels of uncertainty. The approach developed here takes into account only data aggregated over a time period, usually a year. However we know in addition that there is also considerable predictability for urban indicators of the same city across time. Thus, we expect that the future combination of these two elements will yield a procedure to make better predictions of future indicators for specific places with quantified uncertainty. This ability will also allow the detection of exceptional events as statistical anomalies in urban indicators. We hope therefore that our growing quantitative understanding of cities and urban systems throughout the world will provide the basis for the development of a predictive science of cities that will help inform more effective policy in an increasingly urbanized world.

3.4 Materials and Methods

Details about the data can be found in appendix A.

3.4.1 Power-Law Fits

Clauset *et al.* (2009) developed a methodology to estimate the parameters of a power-law fit, and to calculate its associated goodness of fit. The function fitted is the pure power-law

$$P_Y(y) = C\left(\frac{y}{y_{\min}}\right)^{-\tau}, \tag{3.24}$$

where C is the normalization constant. The distributions of homicides analyzed here were fitted using this functional form and we were unable to reject the power-law fit. However, the fit only holds for values of $x \geq x_{\min}$. Following Clauset *et al.* (2009), the estimated p-values were $\widehat{p}_{\text{Col}} = 0.34 \pm 0.05$, $\widehat{p}_{\text{Mex}} = 0.38 \pm 0.05$ and $\widehat{p}_{\text{Bra}} = 0.73 \pm 0.05$, which were not sufficiently small for the power-law distribution to be rejected.

Because we are interested in the regime of small numbers where the number of homicides Y can be zero, we extend equation (3.24) to

$$P_Y(y) = \frac{(y+k)^{-\tau}}{\varsigma(\tau,k)}, \tag{3.25}$$

which converges to equation (3.24) for large y, but with the difference that now y can take any non-negative value.

If we let Y_i, $i = 1,\ldots,n$, be the observed annual number of homicides of each city. Assuming independence, the log-likelihood of the data under equation (3.25) is

$$\begin{aligned} \mathscr{L}(\tau,k) &= \sum_{i=1}^{n} \log\left(\frac{(Y_i+k)^{-\tau}}{\varsigma(\tau,k)}\right) \\ &= -n\log(\varsigma(\tau,k)) - \tau\sum_{i=1}^{n}\log(Y_i+k). \end{aligned} \tag{3.26}$$

A numerical estimation of k and τ by setting $\partial\mathscr{L}/\partial\tau = 0$ and $\partial\mathscr{L}/\partial k = 0$ to maximize the likelihood function, yields

$$\begin{aligned} \widehat{\tau}_{\mathrm{Col}} = 1.864 \quad &; \quad \widehat{k}_{\mathrm{Col}} = 1.904 \quad ; \quad \widehat{p}_{\mathrm{Col}} = 0.6566 \pm 0.005 \\ \widehat{\tau}_{\mathrm{Mex}} = 2.296 \quad &; \quad \widehat{k}_{\mathrm{Mex}} = 1.827 \quad ; \quad \widehat{p}_{\mathrm{Mex}} = 0.4373 \pm 0.005 \\ \widehat{\tau}_{\mathrm{Bra}} = 2.157 \quad &; \quad \widehat{k}_{\mathrm{Bra}} = 1.840 \quad ; \quad \widehat{p}_{\mathrm{Bra}} = 0.0220 \pm 0.005. \end{aligned}$$

A rigorous procedure to estimate these parameters from the data, estimate the error and determine its scaling properties is part of future work.

3.4.2 Lognormal Fits

We test the lognormal distribution as a description of $P_{N|Y}(n|y)$ by standardizing the variables $\ln(N)$ for each given Y, and then showing a normal probability plot (or Q-Q plot) in fig. 3.7. Departures from the lognormal distribution (a normal in logarithmic variables) can be identified by departures from the straight line and are shown, in fig. 3.7, to be both rare and small.

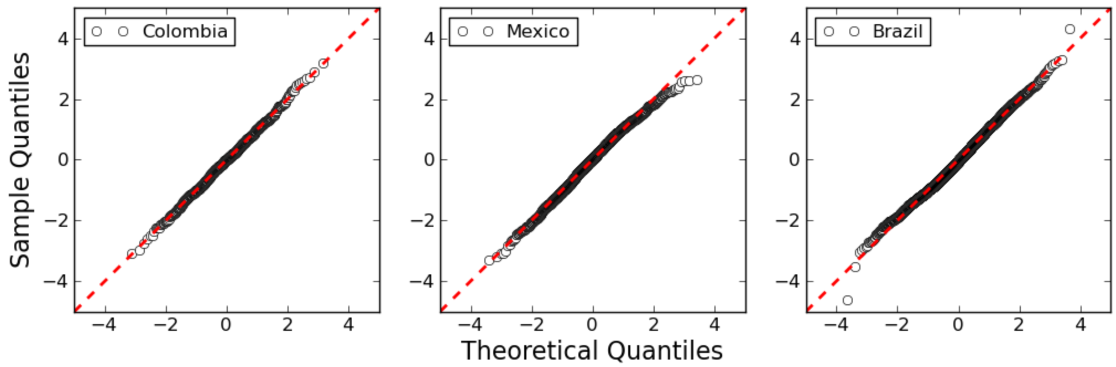


Figure 3.7: Q-Q Plot of the Standardized Log-Variables of the Populations of the Cities for Several Values of Y. This Shows that a Lognormal Distribution is an Excellent Description of $P_{N|Y}(n|y)$, for the Three Nations, Notwithstanding a Number of Small Exceptions at the Extremes (a Perfect Straight Line in the Dots Would Correspond to an Exact Normal Distribution of Log-Populations).

3.4.3 Parameter Estimation of $\mu_y = f(y)$

Maximum likelihood estimations of μ_y for the different values of $y \in \{0, \ldots, 29\}$, for Colombia, Mexico, and Brazil, are shown in fig. 3.6. We constructed a different curve $\widehat{\mu_y}^{(t)}$ for every year of the analysis, and plotted its average $\widehat{\mu_y}$ over the set of annual estimates:

$$\widehat{\mu_y} = \frac{1}{T} \sum_{t \in \mathrm{T}} \widehat{\mu_y}^{(t)}, \tag{3.27}$$

where T the number of years for which we have data, for each nation.

Error bars represent plus and minus one standard deviation about the average.

$$\widehat{\mathrm{err}_y} = \frac{1}{T} \sum_{t \in \mathrm{T}} \left(\widehat{\mu_y}^{(t)} - \widehat{\mu_y} \right)^2. \tag{3.28}$$

The fits were performed using a Levenberg-Marquardt algorithm, which minimizes the sum of least squares of a set of non-linear equations, weighting every point by its error. The function to minimize with respect to vector parameter $\mathbf{p} = (p_0, p_1, p_2)$ is

$$\chi^2(\mathbf{p}) = \left(\frac{\widehat{\mu_y} - f(y; \mathbf{p})}{\widehat{\mathrm{err}_y}} \right)^2, \tag{3.29}$$

where

$$f(y;\mathbf{p}) = \frac{1}{p_0}\ln(y+p_1)+p_2 \tag{3.30}$$
$$= \frac{1}{\beta}\ln(y+r)+\log A. \tag{3.31}$$

3.4.4 Zipf's Law Derivation

Here we give additional details of the calculations leading to section 3.2.5. First, we write $P_N(n)$ in terms of $P_Y(y^*)$ and $P_{N|Y}(n|y^*)$:

$$\begin{aligned} P_N(n) &= \sum_{y^*=r}^{\infty} P(n|y^*)P(y^*) \\ &= \sum_{y^*=r}^{\infty} \frac{1}{n\sqrt{2\pi\sigma_{y^*}^2}}\, \mathrm{e}^{-\frac{\left(\ln n-\mu_{y^*}\right)^2}{2\sigma_{y^*}^2}}\, \frac{(y^*)^{-\tau}}{\varsigma(\tau,k)}. \end{aligned} \tag{3.32}$$

For simplicity of notation, we drop the subscript in σ_{y^*}, and we use the letter y, although it is important to keep in mind that we are implicitly referring to y^*. Replacing the sum with an integral and assuming r is sufficiently small that we can integrate over the whole range of non-negative numbers, we obtain

$$\begin{aligned} P_N(n) &\propto \frac{1}{n}\int_0^{\infty} \frac{1}{y}\exp\left[-\frac{1}{2\sigma^2}\left(\ln n - \ln Ay^{1/\beta}\right)^2 - (\tau-1)\ln y\right]\mathrm{d}y \\ &\propto \frac{1}{n}\int_0^{\infty} \frac{1}{y}\exp\left[-\frac{1}{2\beta^2\sigma^2}\left(\ln^2 y - 2\ln y\left(\beta\ln\frac{n}{A} - \beta^2\sigma^2(\tau-1)\right) + \beta^2\ln^2\frac{n}{A}\right)\right]\mathrm{d}y. \end{aligned} \tag{3.33}$$

We can now complete the square and re-arrange terms to obtain

$$P_N(n) \propto \frac{\exp\left(-(\tau-1)\ln\left(\frac{n}{A}\right)^{\beta} + \frac{\beta^2\sigma^2}{2}(\tau-1)^2\right)}{n}\int_0^{\infty}\frac{1}{y}\exp\left[-\frac{1}{2\beta^2\sigma^2}\left(\ln y - f(n;\theta)\right)^2\right]\mathrm{d}y. \tag{3.34}$$

Now, we see that the term inside the integral is a lognormal distribution, integrated over its entire domain. Consequently, the integral is a constant, regardless of the form of $f(n;\theta)$, where θ represent the parameters A, β, σ and τ. Retaining only terms in n, we obtain

$$\begin{aligned} P_N(n) &\propto \frac{\exp\left(-(\tau-1)\ln n^{\beta}\right)}{n} \\ &\propto n^{-\beta(\tau-1)-1}, \end{aligned} \tag{3.35}$$

from which we finally see that

$$P_N(n) \propto n^{-\alpha}, \tag{3.36}$$

with $\alpha = \beta(\tau-1)+1$, or

$$\beta = \frac{\alpha-1}{\tau-1}. \tag{3.37}$$

This relationship can also be derived in a more straightforward way under the assumptions that i) a power-law distribution for Y (or N) holds and ii) the scaling relationship $Y \propto N^{\beta}$ holds *exactly*. Then, using the fact that $P_N(n) = P_Y(y)\mathrm{d}y/\mathrm{d}n$, we obtain the same relation between exponents. The derivation given above, however, does not assume an exact expression in the form of $Y = f(N)$, but rather a probabilistic relation between N and Y, through the expectation value $\mu_{Y^*} = \ln\left[A(Y^*)^{1/\beta}\right]$.

Figure 3.8 shows the cumulative empirical distributions of city populations.

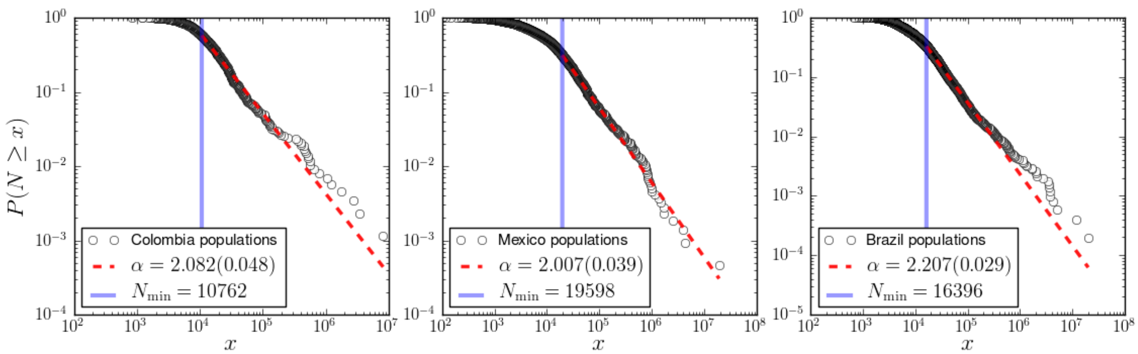


Figure 3.8: Cumulative Normalized Distributions of City Populations in Colombia, Mexico and Brazil (2007) Fitted With Pure-Power-Law Distributions. Best Fits (Dashed Red Line) of the Form $P(N \geq x) = Cx^{-\alpha+1}$ Were Estimated Using the Procedure in Clauset *et al.* (2009) to the Density Function. Not Disregarding the Long-Held Debate about the City-Size Distribution, we Believe the fit to a Power-Law Distribution Stands as a First Approximation Consistent with our Proposed Statistical Framework.

Chapter 4

ARE THERE CONSTRAINTS ON CREATIVE AND INVENTIVE ACTIVITIES IN URBAN AREAS?

A question in public policy is to what extent can the levels of inventive and creative activities be increased in cities to enhance economic development. To reveal the constraints on the processes that determine urban inventive and creative activities, we analyze the lognormal probability distributions that describe the counts of inventors and creative employment conditioned on population size, and the unconditioned distribution of population sizes of U.S. Metropolitan and Micropolitan Statistical Areas. This approach reveals some of the characteristics of the processes that increase or decrease the levels of inventive and creative activity and allows us to estimate the probabilistic constraints on the levels of such activities in urban areas.[1]

4.1 Introduction

Innovation and knowledge creation are by now widely recognized to be the primary drivers of economic growth and development (Weil, 2012). As Charles Jones points out (Jones, 1995, p. 764), knowledge is simply the accumulation of ideas, and ideas are developed by individuals. Urban environments, with their agglomeration of interacting individuals, have historically been the privileged setting for innovation (Hall, 1998). Specifically, innovative, inventive and creative activities, which is to say, innovative, inventive and creative individuals, are concentrated in urban areas (Ó hUallacháin, 1999; Glaeser and Saiz,

[1]This work was done in collaboration with Professors Luís M. A. Bettencourt, Kevin Stolarick, Deborah Strumsky, and José Lobo.

2004; Bettencourt *et al.*, 2007b). Much work has been undertaken to elucidate the factors influencing the clustering of skilled individuals in urban areas (e.g., industry composition, presence of educational institutions, job opportunities, natural and cultural amenities, etc., Glaeser and Saiz, 2004; Berry and Glaeser, 2005; Florida *et al.*, 2008; Glaeser, 2011; Miguélez and Moreno, 2013).

The most common methodological approach for these investigations has been the use of multivariate regressions; these efforts have been successful in identifying many of the determinants affecting the concentration of innovative (or inventive or creative) individuals in urban areas. However, in the same way that the knowledge that human weight is a good predictor for human height does not provide us with information about the probability, or improbability, of observing a 3 meter-tall human being, previous studies about the agglomeration patterns of skilled and talented individuals do not informs us about the probability of observing Boston's creative and inventive levels, or San Jose's, or Zanesville's, or any other urban area. The characterization of the probability distributions describing these levels determines what we mean by the "probabilistic constraints" of creative and inventive activities, and it further informs about some general characteristics of the generative processes that determine those levels. Supplementing our multivariate regressions with explicit calculations of the probabilistic constraints is vital if we want to have an understanding of what is possible, and what can be changed, when implementing public policies.

In our approach, we echo Storper *et al.* (2012), who state that noisiness and stochasticity are an inherent characteristic of urban dynamics. Given the social character of cities, and the presumably large web of influences affecting the locational decisions and intellectual activities of individuals, the processes underlying agglomeration are bound to be stochastic in nature (Curry, 1964). The nature of these processes is largely defined by the manner in which a multiplicity of factors come together, that is, how these factors aggregate, to determine a stochastic outcome. Whether the aggregation is additive or multiplicative, corresponding to separate or interacting co-occurrence of factors, outcomes will display different probability distributions. However, it can be argued that the behavior of a product of random variables is considerably richer than that of a sum of random variables (Redner, 1990) because it models interaction. The logic by which the mathematical operation of multiplication represents interactivity is demonstrated in the functional form of a production function or in the use of a multiplicative term in a regression equation to capture interaction among two independent variables (Aiken and West, 1991). It is precisely the nature of the stochastic aggregation which needs to be understood in order to assess the efficacy of policy interventions that aim to stimulate the attraction and enhancement of creativity in urban areas. By considering the full extent of the aggregated stochasticity we can specify the limits on the variability exhibited by urban areas with regards to their endowments of creative resources. It turns out, how much more creative, or conversely, how uncreative, any one urban area can be is constrained in a systematic way, given its population size. And, although innovative activities are disproportionately located in larger cities, population size itself shows statistical signatures of being constrained as cities become too big.

Here we show that the urban variability in innovative (i.e., creative and inventive) individuals, once population size is controlled for, is well described by a lognormal distribution. Characterizing this distribution informs us as to the limits of such variability – that is, how much more creative or inventive can the very creative or inventive places be. Our results suggests (i) that there is an underlying *multiplicative* stochastic process affecting the levels of urban creative activity, (ii) that the number of determining factors involved is large, and (iii) that there exist constraints on the variance of the logarithmic levels of inventive and creative individuals. These constraints are a manifestation of the underlying generative processes, which in turn is characterized through a probability distribution.

Although the way the results are presented is novel, our conclusions 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 (Ó hUallacháin, 1999; Berry and Glaeser, 2005). We present the evidence in the language of probabilities which we argue advances our understanding of how innovation in urban areas operates, and opens new avenues of research as well.

4.2 Research Design

In the absence of a clear, and measurable, definition of innovation we believe that a broad statistical approach to proxy quantities is essential. Such an analysis takes as its primary quantities, not the covariances between variables, but rather their statistical distributions. The main justification for focusing on statistical distributions is that uncovering them can provide insights into the underlying processes generating observables (Frank, 2009; Frank and Smith, 2011; Sornette, 2012), as well as the likelihood of observing any given value. We investigate the probabilistic nature of the processes concentrating creative activities in urban areas essentially in two ways. First, in the cross-sectional regression

$$\log(Y_i) = \log(Y_0) \ + \ \beta \log(N_i) \ + \begin{array}{c}\text{Aggregated effect}\\ \text{of all other factors}\end{array} \quad , \tag{4.1}$$

where Y_i represents the counts of skilled individuals and N_i the total population size of the i-th urban area, we allow the error term to aggregate the effect of all the other factors affecting the level of skilled individuals Y_i. This is in contrast to the typical approach which disaggregates the error term into several control, instrumental, and explanatory variables (in order to minimize the residual variability). Instead, we characterize the probability distribution that describes these error terms, which is equivalent to studying the shape of the probability function $P(Y|N)$ (recall section 1.2). Second, we characterize the probability distribution of population size $P(N)$. In this way, using Bayes' rule, the joint probability distribution $P(Y,N)$ is estimated. We introduced this approach in chapter 3, and as was mentioned in that chapter, it has already been pursued in similar contexts recently (Bettencourt *et al.*, 2010; Gomez-Lievano *et al.*, 2012), but here we analyze in more detail what the shape of the distribution suggests about the constraints acting on our observables. These constraints in turn are indicative of features of the processes determining the inventive and creative profiles of urban areas that act across the whole urban system.

The decision to use population size N as the conditioning factor is widely adopted in the literature, but is particularly relevant in our context since our proxy measures are counts of people. For the sake of clarity, it is important to recall what our results in chapter 2 were. Although the results applied to a simple null model of a city, two main points were made that could apply to more complex situations. First, that power-law relations between measures of total aggregate output and population size N were not spurious, and in fact, emerged precisely because population sizes were not large enough to guarantee convergence of the law of large numbers. And second, because of the latter, that per capita measures should be avoided since their correct interpretation assumes specifically such convergence. For purposes of our investigation, characterizing the probability distribution of residuals, using aggregate measures of creative or inventive activity (e.g., counts of inventors), as opposed to per capita measures (i.e., inventors per capita), makes no difference since the choice only affects the β coefficient. Conceptual and probabilistic considerations make the use of per capita measures far from an innocent choice, requiring further assumptions whose consideration were analyzed in chapter 2.

4.2.1 Data and Definitions

We will refer to the "creative endowments" of a city as its creative class workers (Florida, 2004) and inventors (i.e., authors of patents), and we will consider their counts as proxies for the creative and inventive activity in a city, although we will analyze each measure separately and in parallel.

In this study, our spatial unit of analysis are the $m = 938$ U.S. metropolitan and micropolitan statistical areas.[2] The number of such areas can vary, depending on the current definitions, and on whether one omits or not Puerto Rican areas. The analysis presented here is not sensitive to these changes. Here we use the 2010 definitions, and we apply the same definition for the years 2008 and 2009. Metro and micro areas are defined as statistical geographic entities, usually a set of counties, consisting of a core area (or *urbanized* area as defined by the U.S. Census Bureau) and the adjacent counties with strong commuting ties. This specification is an attempt to define cities as socioeconomic entities and integrated labor markets. The population of the core for metro areas is at least $50,000$ and for micro areas is at least $10,000$ (but less than $50,000$). This means that our data is left-censored, such that population sizes below $10,000$ are excluded. For each urban area we use the population size estimates from the US Census Bureau.

To construct creative occupations employment numbers, defined in (Florida, 2004, Appx. A), we use 2010 employment data from the U.S. Department of Labor's Occupational Employment Statistics (OES) that is 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 was obtained from coding inventor's addresses obtained from the U.S. Patent and Trademark Office (USPTO).[3] 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

[2]Micro and metro areas are collectively referred to as "core based statistical areas" (CBSA), but we refer hereafter to the metropolitan and micropolitan areas as urban areas, or simply as cities, interchangeably. Definitions and more information can be found in `http://www.census.gov/population/metro/`.

[3]`http://www.uspto.gov/`.

locate them geographically. Details about the algorithm used can be found in Marx *et al.* (2009).

4.2.2 Estimations of Probability Functions

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. In the present discussion the former represents counts of either creative workers or inventors, while the latter represents population size. The constraints on the levels of creative and inventive activity are inferred directly from the data, and quantified through the estimation of the cross-sectional probability distributions describing the corresponding measures of creative class employment and inventor counts. Since both measures 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 specify $P(Y = y|N = n)$, 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 of having the stated population size $P(N = n)$. We show in appendix G that knowing the marginal distribution $P(Y)$ helps us estimate $P(N)$.

We will relax the condition that Y and N must take integer values, and instead consider them to be continuous. In the case of inventors, we will take the 3-year average from 2008 to 2010. Since our counting of inventors in each year depends upon the existence of patent applications, our numbers are subject to interannual fluctuations. Using a 3-year average reduces this variation. In the case of creative employment, the use of a continuous approximation is valid since more than 99 percent of the count numbers are larger than $1,000$ and span a range of more than three orders of magnitude. Under these conditions the granularity of the data is not so evident, as $P(Y=y) \to 0$. The approximation is even more valid for population size. Hence, we will be estimating probability *density* functions (pdfs), as opposed to probability *mass* functions, which we will denote by $p_{Y|N}(y\,|\,n\,;\,\theta_n)$, $p_N(n)$, $p_Y(y)$, where we have made 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). In practice, $P(Y|N)$ is estimated using logarithmic binning, i.e., $P(Y|n_j \leq N < n_{j+1})$, 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 in urban metrics for cities of comparable population sizes. Conversely, the marginal distribution characterizes the sizes of the concentrations of people across the whole urban system. Note that by estimating $P(Y|N)$ and $P(N)$ we are also estimating the joint probability function of Y and N, since $P(Y,N) = P(Y|N)P(N)$.

To understand the constraints on population size we also consider urban population's rates of growth, relative to the whole US population growth rate. For the analysis of urban population growth rates, we took the decennial census records between 1800 and 2010 from the U.S. Bureau of the Census. Based on the current definitions of metropolitan statistical areas we constructed the data for MSA populations going as far back in time as possible.[4]

4.2.3 Interpretations of Probability Distributions

Strictly speaking, there is an infinite number of different stochastic processes that can generate any particular probability distribution describing a generic random variable X. Thus, observing a particular statistical distribution does not immediately and uniquely de-

[4] `http://www.census.gov/population/www/censusdata/hiscendata.html`.

termine 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 $X_1, X_2, \ldots$. Hence, some limit theorems may apply, and thus we expect to observe some of the limiting distributions.

Here 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 X, a situation which can be mathematically expressed as $X = \sum_i X_i$. If this is the case, we expect the random variable X to be normally distributed by invoking the Central Limit Theorem (CLT). In contrast, if factors are acting interactively, which can be represented as $X = \prod_i X_i$, we expect the outcome X to be *lognormally* distributed, which can be understood by using the CLT once we apply logarithms to both sides of the equation. The standard assumptions behind CLT are that X_i must be independent and identically distributed, and have a finite variance. 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 X_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.

4.2.4 Visualization of Probability Distributions

We will visualize the conditional probability density function $p_{Y|N}(y \mid n)$ by plotting histograms of the data. This is customary for visualizing empirical distributions in general, but it becomes less useful when the tails of the distributions are heavy and too noisy. This is precisely the situation for urban population sizes, in which the largest cities are very few and have extreme large values. In this case, distributions are best visualized using the cumulative distribution, or the complementary cumulative (or "countercumulative"), which are more robust to the noise in the tails. Thus, we will plot the marginal probability $p_N(n)$ through its complementary cumulative function

$$P(N \geq n) = \int_n^\infty p_N(n')\mathrm{d}n'. \tag{4.2}$$

All the cumulative plots will be shown in double logarithmic scales.

4.2.5 Limitations of Our Analysis

Any study that analyzes processes will be necessarily incomplete without studies of change through time. Therefore, our cross-sectional study of the US urban system is incomplete and must be complemented with future studies that take into account time. We have relied on previous studies to inform our understanding of how the quantities we measure and estimate change through time. Batty (2006) reports that the rank-size rule (Zipf's Law) of population sizes is not universal, and concludes that Gibrat's Law of proportionate growth needs to be revised. We give further empirical support of this in our analysis. From Bettencourt *et al.* (2010), we highlight the fact that urban aggregate measures, in general, are highly correlated in time. This is relevant to our study since it indicates that measures of cross-sectional variation are the result of cumulative effects over time.

4.3 Results

Our specific questions are: How many creatives and inventors should we expect there to be in a city (i.e., estimating $E(Y|N)$)? What is the population size distribution, and how is population size constrained (i.e., estimating $P(N)$)? How do cities of comparable population size differ in their creative endowments (i.e., estimating $P(Y|N)$)? The answers to these three questions characterize statistically the urban system's creative endowments (in

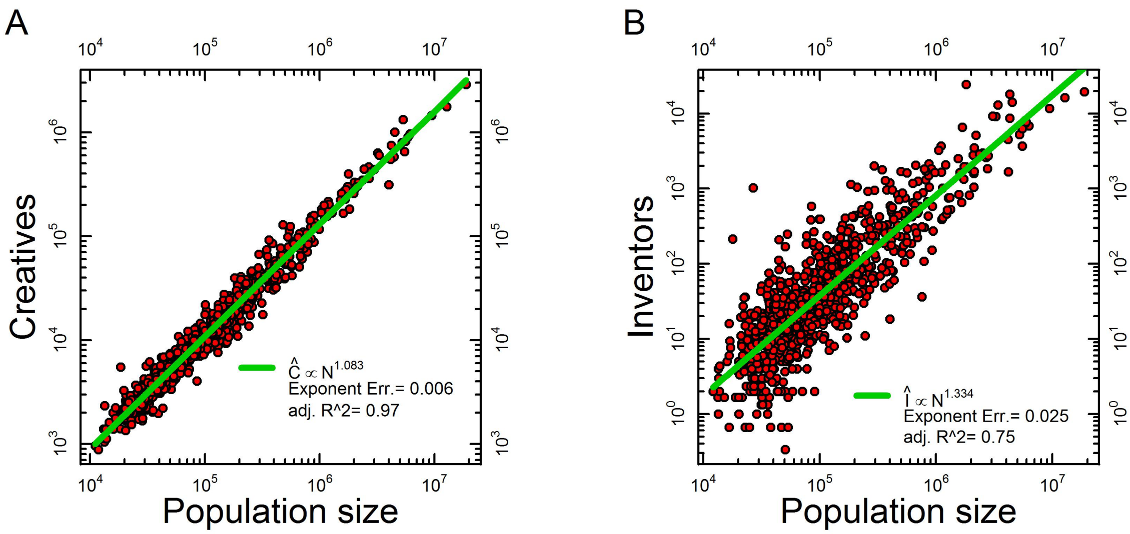


Figure 4.1: Population Size Scaling of the Counts of Creative Employment C, and Inventors I, in U.S. Metro and Micropolitan Areas. In Both Plots, Each Point Represents a Different Metropolitan or Micropolitan Area. The Data on Creative Employment Correspond to the Year 2010, and for Inventors it is the 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 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).

the sense defined above) and reveal its constraints. We start answering the posed questions by showing the general relationship between the urban indicators y and n in a scatter plot (fig. 4.1). This motivates the investigation of population size alone, in which we determine the probability density function $p_N(n)$ for moving in the *horizontal* direction in the scatter plot (fig. 4.2). Once this has been determined, we continue with the analysis of the probability density function that describes the residual *vertical* variation, $p_{Y|N}(y\,|\,n)$ (fig. 4.3). Since population size sets the baseline for how most of the processes that agglomerate creative and inventive individuals operate in cities, it is the characteristics of this conditional distribution that we focus most of our attention on.

4.3.1 The Role of Population Size and Its Constraints

More than 70 percent of the cross-sectional variation in creative employment and inventor counts is explained by population size alone; the remaining variation, less than 30 percent, is determined by other factors, many of which are the focus of a vast literature in urban economics, economic geography and regional science. We single out population size from among all possible explanatory factors as the conditioning variable, because of its large explanatory power, its customary use as a control variable, and because phenomenologically it takes precedence over other urban characteristics as a necessary condition for social systems to work. The importance of population size as a determinant of socioeconomic life has long been noted by archaeologists, anthropologists, sociologists and economists; we refer the reader to some of the recent literature and references therein, e.g., Quigley (1998); Bettencourt *et al.* (2007a,b); Jones and Romer (2010).

Figure 4.1 plots creative employment and inventors counts against population size over all metro and micropolitan statistical areas, together with a linear OLS regression of the logarithmic variables (green solid line). The regression tests the hypothesis that the expected value is given by

$$E(Y|N) = Y_0 N^{\beta}, \tag{4.3}$$

where Y_0 and β are coefficients whose estimates are shown in fig. 4.1.

The observation that $\hat{y}_i \propto n_i^{\hat{\beta}}$, where $\hat{\beta}$ is the estimate of the exponent of this power relation (4.3), is indeed a good predictor of how these quantities scale with population size serves as the starting motivation for our statistical analysis. The fact that $\beta > 1$ means that larger urban areas are skill abundant, and one may be inclined to think that the key to increasing creative and inventive activities in cities is to increase their population size. However, when we observe how many cities there are of different sizes through the estimation of $p_N(n)$, we observe that becoming increasingly larger becomes increasingly difficult.

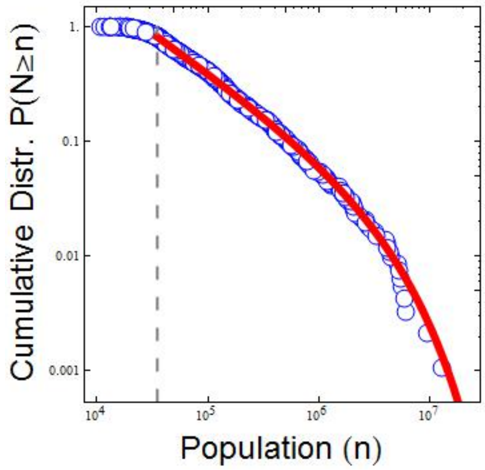


Figure 4.2: Complementary Cumulative Distribution of Population Sizes. A Cutoff in the Largest Population Sizes is Present, Suggesting an Accumulated Effect from Constraints to Growth. For Fitting the Density Function (4.4) of $p_N(n)$, we Used the Estimated Parameters of the Density $p_Y(y)$ for Creative Employment Counts (See appendix G). 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 G for Goodness-of-Fit Tests.

Figure 4.2 shows the empirical complementary cumulative distribution of population sizes. The function

$$p_N(n;\alpha,\beta,\nu,n_{\min}) = C\frac{\mathrm{e}^{-\left(\frac{n}{\nu}\right)^{\beta}}}{\left(\frac{n}{\nu}\right)^{\alpha}}, \quad n \geq n_{\min}, \tag{4.4}$$

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 (4.3). 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). We refer the reader to the appendix G for the details about the estimation of the parameters and the derivation behind the functional form in equation (4.4).

The rationale behind equation (4.4) 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" 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). Second, and despite the previous reasoning, 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 an exponential decay (Newman, 2005).

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 fig. 4.2, however, only the middle range of US city sizes displays this scale-free phenomenon, and deviations from it in the right tail suggest that scale becomes significant for large populations. Estimates of the population sizes at which these regime transitions occur are given by $\widehat{n_{\min}} \approx 35,000$, from which the Pareto behavior starts, and $\widehat{\nu} \approx 11,000,000$, in which the Pareto behavior ends. The cities with population sizes larger than $\widehat{\nu}$ are the New York MSA with approximately 19 million and the Los Angeles MSA with approximately 13 million inhabitants. These boundaries are not absolute, and are rather smooth transitions. The important result in this section is that we have estimated a size scale around which stronger constraints to population growth appear.

[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.

[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.

Note that we are less interested in evaluating the exact shape of $p_N(n)$ and more interested in what the distribution tells us about the constraints on population size. The specification given in equation (4.4), and its superior fit over other specifications (see appendix G), suggest that there is in fact a scale above which 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 the cutoff is due to a geographic underspecification of these regions.

In consequence, we see that the U.S. urban system as a whole has three population-size regimes, with two characteristic population scales separating them. Since the number of creatives and inventors positively correlates with population size, their marginal distributions mirror that of the population when looked across all cities (for a similar analysis about patents and population size distributions see Ó hUallacháin, 1999). If a Pareto distribution is taken as the null description of city population sizes (Batty, 2008), then we identify deviations from it in the U.S. system of metropolitan and micropolitan areas on the left and right tails of the distribution. Especially interesting is the deviation in the upper tail, in which big cities are smaller than what a Pareto distribution would predict (see also Black and Henderson, 2003 for an econometric analysis of such deviation). This observation suggests different population growth dynamics appear after a certain population size. Data on relative grow rates of metropolitan cities supports this hypothesis (see appendix G, fig. G.2). A cutoff in the right tail of the population size distribution effectively reveals a constraint on big population sizes.

The distribution of population sizes has a long history that goes back to the first half of the twentieth century when it was discovered the remarkable regularity that the sizes of the largest cities displayed a linear relationship when plotted against their rank in log-log plot. The debate has more recently been on whether this distribution is best described by a log-normal, a Pareto, or different mixtures of both distributions in which the body is more lognormal-like with a Pareto tail (Ioannides and Skouras, 2013; Malevergne *et al.*, 2011; Giesen *et al.*, 2010).

These previous studies, however, use as their main units of analysis the U.S. Census Places[7]. For such units of analysis, these studies provide good reasons to believe that the body of the distribution might be lognormal and that the tail is Pareto. Since such urban unit definitions are not motivated by economics questions, the resulting city size distribution does not convey information about how cities operate as socioeconomic networks. Metropolitan and micropolitan areas, on the contrary, are more germane units for urban economic analyses. Given this, the observed deviations from a Pareto distribution in the upper-right tail of metro and micro populations sizes can be interpreted to arise, from this point of view, from the socio-economic processes that constrain or enhance the growth of cities.

Even though we have not studied the factors that explain this population size cutoff, it is not unreasonable to think that there are different growth dynamics in the scale of the largest cities within an urban system. In fact, the general statistical pattern and its interpretation regarding growth dynamics has been recognized before (e.g. Black and Henderson, 2003; Duranton and Puga, 2004; Rosenthal and Strange, 2004, 2006). What is new in our approach, is that we have identified such effect through *other* urban metrics. This has made the statistical estimation procedures to perform better, which has allowed us to better estimate the scale at which these constraint act, by estimating a cutoff *scale* (see appendix G). It is important to emphasize that the effect of the cutoff is relative, and that the largest cities are still growing with the proportion of creative employment and inventors to total population still scaling regularly across all sizes.

[7]The U.S. Census Bureau defines a place as concentrations of population that have a name and are independently and locally recognized.

4.3.2 The Distribution of Creative and Inventive Activities for Cities of Comparable Population Size

Figure 4.1 shows that the average of creative employment and inventors counts varies regularly as a power of population size, characterized by equation (4.3). 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 4.1 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

Figure 4.3 plots the histogram of the transformed variables

$$z = \frac{\ln\left(y_{|B_j}\right) - \widehat{\mu}_j}{\widehat{\sigma}_j}, \tag{4.5}$$

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 $y_{|B_j}$:

$$\widehat{\mu}_j = \frac{1}{\left|B_j\right|} \sum_{n_j \in B_j} \ln\left(y_j\right), \tag{4.6}$$

and

$$\widehat{\sigma}_j = \sqrt{\frac{1}{\left|B_j\right| - 1} \sum_{n_j \in B_j} \left(\ln\left(y_j\right) - \widehat{\mu}_j\right)^2}. \tag{4.7}$$

In equations (4.6) and (4.7), $\left|B_j\right|$ denotes the number of observed cities in bin B_j.

The quantity z for all variables is well fitted by a standard normal distribution (fig. 4.3), which means that the untransformed numbers y_i are lognormally distributed, conditioned

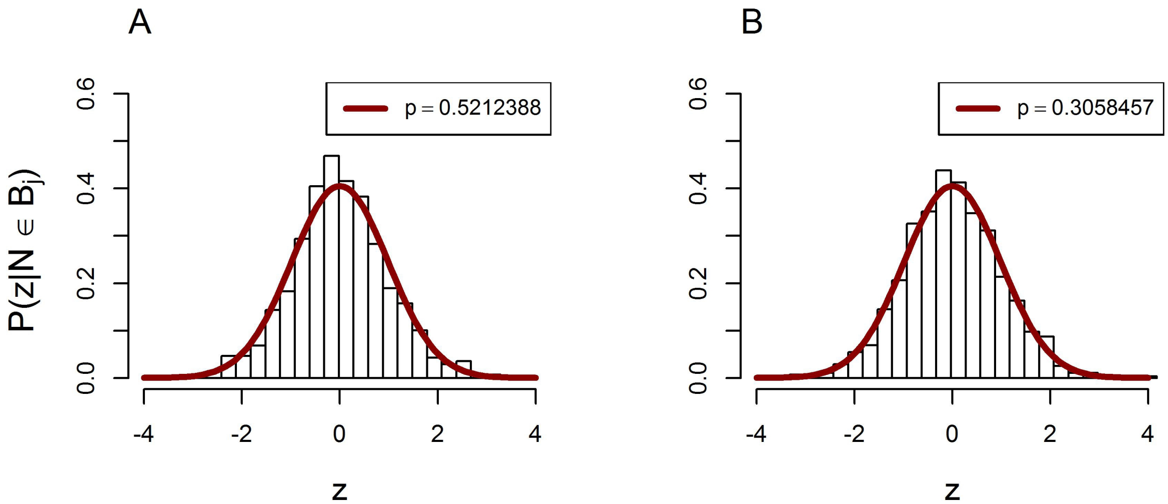


Figure 4.3: Histograms of Creative Individual Counts Conditional on Population Size. We Fit a Standard Normal Distribution to the Normalized Frequency Histograms of the Standardized Logarithmic Counts of Creatives (A) and Inventors (B). For Both Plots the Counts Were Transformed $y_{|B_j} \to z = \frac{\ln(y_{|B_j}) - \widehat{\mu_j}}{\widehat{\sigma_j}}$, 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 (~ 25) and the Total Number of Bins (~ 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 G for More Details on Goodness-of-Fit Tests.

on population size.[8] The density of the random variable Y can thus be written as:

$$p_{Y|N}(y \,|\, n\,;\, \mu_n, \sigma_n) = \frac{1}{y\sqrt{2\pi\sigma_n^2}} \mathrm{e}^{-\frac{1}{2\sigma_n^2}(\ln(y)-\mu_n)^2}. \tag{4.8}$$

We conclude that a lognormal distribution provides a good description of the number of creative class workers and inventors for all cities, given the population size dependence of the distribution parameters. Goodness-of-fit tests are presented in the appendix G.

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). Thus, it is not a

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

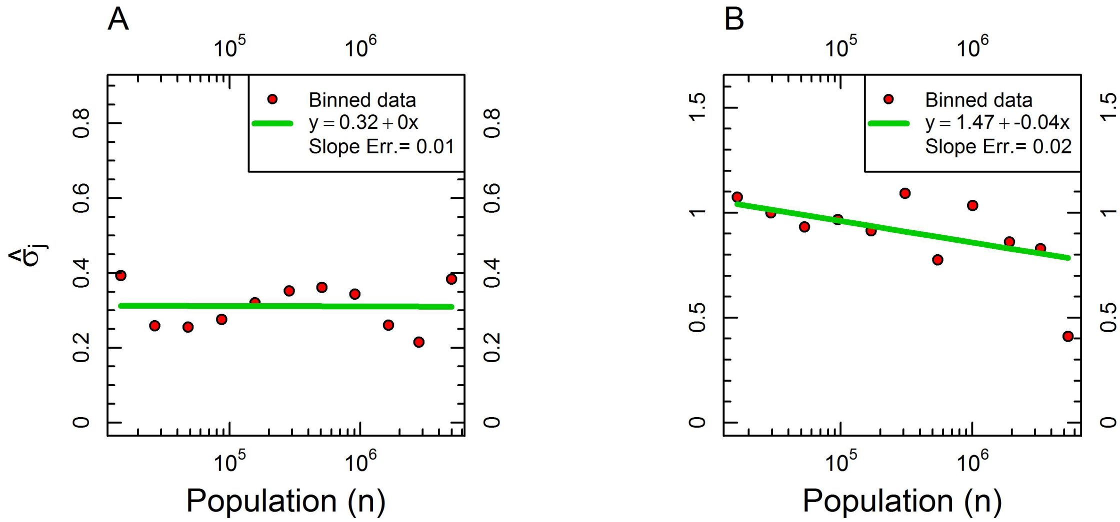


Figure 4.4: Population Size Dependence of the Estimates of the Lognormal Parameters of Each Bin. Creatives (A) and Inventors (B) Are Shown Here to Differ in How the Variance of the Logarithmic Counts $\widehat{\sigma_n^2} = \sum(\ln Y - \widehat{\mu_n})^2/(m-1)$ Depends on Size, and the Level. Although the Variance is Much Higher for Inventors, there is a Weak Tendency to Decrease with Population Size. This May Be an Artifact of the Last Point, Which only Has Few Data Points (See Text).

surprise to find the distribution of creatives and inventors to be lognormal. The importance of this is that it can be quantified, and used to assess the viability of policy goals.

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 4.4 plots the n-dependence of the parameter σ_n of those bins.

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. Since the standard deviation of a lognormal distributed random variable is proportional to the mean, this is equivalent to saying that the standard deviation s of the variable Y scales with population size as

$$s(Y|N) \propto N^{\beta_\sigma}, \tag{4.9}$$

where β_σ can be a constant as for creatives (fig. 4.4A), or a weakly varying function of N as for inventors (fig. 4.4B). Since the levels of creative and inventive activities are strongly correlated from one year to the next, 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 fig. 4.1. 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 4.1 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 $y \notin [\mathrm{e}^{\mu(n)-z\sigma(n)}, \mathrm{e}^{\mu(n)+z\sigma(n)}]$, where the value of $z = 3.07$, such that $1 - \Phi(z) < 1/938 \approx .00107$, where $\Phi(\cdot)$ is the cumulative normal standard distribution.

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 4.1 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.

Table 4.1: U.S. Urban Areas, Sorted by Population Size, Having Counts, in Inventors or Creatives, That Are Outside a $z = 3.07$ *sigma* Interval Around the Log-Mean, i.e., $y \notin [e^{\mu(n)-z\sigma(n)}, e^{\mu(n)+z\sigma(n)}]$. 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

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.

4.3.3 Multiplicative Random Processes

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} 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.

From equation (4.1), we conclude that

$$\begin{array}{l}\text{Aggregated effect} \\ \text{of all other factors}\end{array} = \sum_{j=1}^{m} f_j(X_{i,j}), \tag{4.10}$$

where the $X_{i,j}$ is the j-th factor influencing the creative endowments of city i, and f_j is a function of this factor. Our observation that these deviations are (log)normally distributed suggests that the number of factors m is large enough so as to render the Central Limit Theorem applicable. These empirical observations and their meaning can be thus summarized in the following equation:

$$y_i = Y_0\, n_i^{\beta} \left(\prod_{j=1}^{m} g_j(X_{i,j}) \right), \tag{4.11}$$

where $g(x) \equiv \mathrm{e}^{f(x)}$. The fact itself that equation (4.11) is multiplicative is indicative that these factor influence the variable Y in an *interactive* way. This suggests that the level of creative and inventive activities in an urban area is a structural outcome. The parameters of the distributions impose further constraints: the mean and standard deviation of these quantities scale as power functions of population size.

Lognormal distributions 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* ..."). For instance, the right regional amenities, job opportunities, partnerships, etc., all have to act in consonance in order to increase its creative endowments (from within and/or from outside). 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.

4.4 Conclusion

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 mechanisms underlie the agglomeration of creative and inventive individuals. It is likely, however, that the number of mechanisms is large. Correspondingly, we have focused our attention into analyzing 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".

As nations draw more heavily on their capacity to produce knowledge to foster economic growth, and as their 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 become central. Here we have delineated, in precise quantitative terms, the probabilistic landscape in which cities embedded in a larger system can be found.

The reason probabilities take a central role in this investigation is because 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. This, effectively, can be modeled in a stochastic way. The study of statistical fluctuations is necessarily richer than the study of averages. A system in equilibrium has its forces in balance, and consequently the study of averages is unable to reveal those forces. The forces are manifested, rather, when the system fluctuates from its equilibrium or average state. By studying these fluctuations, and how they change with population size, a simple characterization of the statistics and behavior of cities emerges.

As a baseline to understand the general statistical patterns innovative activities in urban areas, we studied the constraints on population size. We found that, at least for the case of

U.S. Micropolitan and Metropolitan areas in recent years, increases in population become more difficult once a city enters a stage of large population sizes. Extending this study to other countries constitutes future work.

Overall, the number of creative and inventive individuals increase on average superlinearly with population size, but the deviations from the mean behavior are mediated through some constrained multiplicative random process. This reveals that the array of different conditions interact, and thus the levels of urban creative and inventive activity are either amplified or suppressed, and this emphasizes the need to have structural views of the functioning of cities.

Although we have focused on counts of creative employment and inventors as sources of creative and inventive activities to understand urban wealth creation, the methodology presented here stands as a general framework to study cities and urban systems.

A clear statistical description of cities across an urban system had been missing as a way to understand innovation in urban systems. A detailed stochastic model that reproduces the statistics found here is part of ongoing work. 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.[9]

[9]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.

Chapter 5

COMPUTATIONAL MODEL OF SKILLS DIVERSIFICATION AND URBAN DIFFERENTIATION

In the previous chapters we have seen how aggregate measures of urban output are approximately described by lognormal distributions, when conditioned over cities of the same population size. Alves et al. *(2013a,b, 2014); Mantovani* et al. *(2013) have recently reported additional independent support of this result. They have found that lognormals are present in crimes other than homicides, in the number of voters in electoral processes, and other urban indicators in Brazil such as child labor, illiteracy, income, sanitation and unemployment. This is not a trivial fact, since the lognormal distribution is not a limiting distribution arising from sums of random variables. The question is thus, what mechanism or mechanisms of production generate aggregate urban output with lognormal-like distributions? In this chapter we analyze such question with a computational model of production using the framework proposed by Hidalgo and Hausmann (2009) and Hausmann and Hidalgo (2011). We find that the proposed model robustly generates distributions that are approximately lognormal, and that increases in the complexity of skills in a city increases its aggregate production exponentially.*[1]

5.1 Introduction

Recent studies indicate that cities, regions, and countries gain knowledge by creating and attracting skilled individuals (Muneepeerakul *et al.*, 2013; Neffke and Henning, 2013; Neffke *et al.*, 2013). Florida (1995) dubbed this phenomenon the *learning region*. This

[1]This work was done in collaboration with Professor Rachata Muneepeerakul.

process of acquiring and accumulating skills has been shown to bolster the evolution of economies (Glaeser and Saiz, 2004), from less to more complex (Klimek *et al.*, 2012), but this process is constrained (Hidalgo *et al.*, 2007; Muneepeerakul *et al.*, 2013; Bahar *et al.*, 2014). And the constraints come from the fact that new skills have to integrate with the capabilities already present in the economy. Florida (1995, p. 534) describes this situation as an interplay between different types of *infrastructures*:

> Learning regions provide the crucial inputs required for knowledge-intensive economic organization to flourish: a manufacturing infrastructure of interconnected vendors and suppliers; a human infrastructure that can produce knowledge workers, facilitates the development of a team orientation, and which is organized around life-long learning; a physical and communication infrastructure which facilitates and supports constant sharing of information, electronic exchange of data and information, just-in-time delivery of goods and services, and integration into the global economy; and capital allocation and industrial governance systems attuned to the needs of knowledge-intensive organizations.

It is indeed intuitive that a functioning infrastructure must be in place in order for a skill to be productive. This type of constraint, in which an outcome depends on the success of many other factors acting in conjunction (as opposed to separately), is known to generate lognormal distributions, and fat-tailed distributions more generally (classic references are Montroll and Shlesinger, 1982; Redner, 1990; Mitzenmacher, 2003). Hence, the multiplicative process we invoked in chapter 2 to generate the productivity of individuals might be behind the productivity at the level of the whole city instead. But how exactly? Could the view expressed by Florida in the above passage be formalized to provide an explanation of why we see aggregate urban output lognormally distributed in cities of the same population size? We analyze this question from a computational point of view, and formalize these interdependencies between skills in a mathematical model of urban production processes.

Our results suggest that urban areas not only need a diversity of skills to enhance productivity, but also the "right" mix of skills. In the context of our model, production processes are characterized by Leontief production functions of the inputs (comparisons with other production functions can be found in appendix F), which introduce a multiplicative mechanism of production. I will explore and expand the recent theoretical framework of Hidalgo and Hausmann (2009) and Hausmann and Hidalgo (2011), originally proposed to explain the economic development of countries in terms of their diversification, to understand this *constrained multiplicative process* of skills accumulation. This study will address how variability in the aggregate output of cities results from the variability in the diversity of skills at the level of individuals.

5.2 Conceptual Framework

Our question of interest is why urban measures of output are lognormally distributed for cities of the same population size. We base our study on the works of Hidalgo *et al.* (2007); Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011) who model the structure of the bipartite network describing *countries* and the *products* they export. We choose these works because they address questions about the complexity and structure of economies directly using network theory, and thus offer a starting point to study multiplicative processes at systemic levels.

Despite the premise that effects acting in conjunction in the production of an output can be modeled by stochastic multiplicative processes that generate lognormal distributions, it is important to say that there is not a sharp divide between "multiplicative processes" and "additive processes". Cities are noisy environments with several effects affecting each other in a variety of ways. Hence, our aim is not to reproduce exact distributions (e.g., normal or lognormal), but instead to study how multiplicative mechanisms can arise in cities.

The transition between additive and multiplicative processes can be understood by noting that if the multiplicative effects are weak, the process can be approximated by an additive process. For instance, if the effect i is modeled as $(1+\varepsilon_i)$ where ε_i is a random noise, the output

$$y = (1+\varepsilon_1)(1+\varepsilon_2)\cdots(1+\varepsilon_n) \tag{5.1}$$

can be approximated by

$$y \approx 1+\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_n, \tag{5.2}$$

if $\varepsilon_i \ll 1$, by neglecting high order terms.

5.2.1 Production in Cities

We will use cities as our geographic unit of analysis, and we will use the word "products" as a way to refer in general to different observable outputs resulting from production processes in cities.

We expand the model proposed in Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011) by

1. Determining not only the presence or absence of capabilities/skills in a city, but also their *quantity*.
2. Modeling the evolution in time of these endowments.

The model is formally written as (analogous to Eq. (7) in Hausmann and Hidalgo, 2011):

$$\mathbf{X^t} = \mathbf{C^t} \odot \mathbf{P}. \tag{5.3}$$

In general, the matrix $\mathbf{C^t}$ is a multi-dimensional representation of cities and their skills. It is a matrix that can be arbitrarily detailed, having for example the information of how many skills there are in what sector and industry of the economy, with a given level of education, and so on. This matrix, in principle, is time-dependent, since cities acquire and lose skills by general processes of birth, death and migration, which is why we use the superscript $\mathbf{t}$. In contrast, the matrix $\mathbf{P}$ is fixed and constant in time, and represents the production requirements of all products in terms of skills. The operator $\odot$ defines how the endowments of cities are translated into actual products.

Here, we will assume that $\mathbf{X^t} = (X_{cp})_{N_c \times N_p}$ is a $N_c \times N_p$ city-product matrix where the element X_{cp} represents the production value, or urban output, of city c in product p. The matrices $\mathbf{C^t} = (C_{ca})_{N_c \times N_a}$ and $\mathbf{P} = (P_{ap})_{N_a \times N_p}$ are two-dimensional matrices. The matrix $\mathbf{C^t}$ represents the matrix of cities and their skills (of which there are N_a number of them, and which we index by the letter a). The matrix $\mathbf{P}$ is a $N_a \times N_p$ matrix whose columns define products as particular sets of these N_a skills that are required to produce them. In this conceptualization of the world, differences in total urban output between cities will arise from differences in their skills set. Total urban output is mathematically defined as

$$y_c = \sum_{p=1}^{N_p} X_{cp}. \tag{5.4}$$

The resemblance of equation (5.4) with equation (2.5) comes from the fact that the definition of aggregate output means that it is a sum of terms. Equation (5.4), however, is a sum over products produced at a higher level of organization, and not over what individuals produce, like in the null model presented in chapter 2. Also, these terms in equation (5.4) are not identically nor independently distributed. All these differences generate a some patterns of urban differentiation that will be analyzed below.

The $\odot$ operator in equation (5.3) is a production operator that determines how much of each product each city produces, given the skills that it possesses. As an example, and without loss of generality, consider a product p, represented by the vector $\vec{p}$ (taken as a column of matrix $\mathbf{P}$), and defined as requiring three specific capabilities $a = i,\ j$, and k.

Mathematically, the i-th, j-th, and k-th elements of the p-th column vector of $\mathbf{P}$ are α_i, α_j and α_k, respectively, and the rest of elements P_{ap} for $a \notin \{i, j, k\}$ are 0. Now, suppose that a city $\vec{c}$ (a row of matrix $\mathbf{C}$) is endowed with $C_{c,i}$, $C_{c,j}$ and $C_{c,k}$, specifying a measure of the quantity of those specific skills available in the city. These three values represent the input production factors to produce p by city c. We thus have that the total output of this particular product is determined by a production function

$$X_{cp} = F_{\odot}\left(C_{ci}, C_{cj}, C_{ck}; \alpha_i, \alpha_j, \alpha_k\right). \tag{5.5}$$

For completeness, we will consider three options for $\odot$, but we will focus mostly on Option 1, and leave some of the computational results for the other two options for appendix F:

Option 1: A Leontief production function: $F_{\odot}\left(C_{ci}, C_{cj}, C_{ck}\right) = \min\left\{\alpha_i C_{ci}, \alpha_j C_{cj}, \alpha_k C_{ck}\right\}$.

Option 2: A Cobb-Douglas production function: $F_{\odot}\left(C_{ci}, C_{cj}, C_{ck}\right) = \left(C_{ci}^{\alpha_i}\, C_{cj}^{\alpha_j}\, C_{ck}^{\alpha_k}\right)^{\frac{1}{\alpha_i+\alpha_j+\alpha_k}}$.

Option 3: A linear production function: $F_{\odot}\left(C_{ci}, C_{cj}, C_{ck}\right) = \frac{\alpha_i C_{ci}+\alpha_j C_{cj}+\alpha_k C_{ck}}{\alpha_i+\alpha_j+\alpha_k}$.

Here, the coefficients $(\alpha_i, \alpha_j, \alpha_k)$ are just constants, whose precise interpretation changes depending on the production function in consideration. Roughly, they quantify the weight (e.g., productivity) of each factor in the production process. These coefficients are, in principle, introduced through the definition of $\mathbf{P}$. For simplicity, we will take $\alpha_i = \alpha_j = \alpha_k = 1$. The above production functions will apply not only to three factors of production, but to all that are defined by the matrix $\mathbf{P}$.

5.2.2 Limitations

One of the limitations of the model is that it assumes that skills in cities are infinitely available. In other words, the skill endowment of a city represented by C_{ca} is shared by all the processes of production within the city and is not reduced by the number of products requiring the skill. Hence, skills in our model, once acquired, are an unlimited resource. This is unrealistic since a skilled workforce is a scarce and limited resource, which precisely explains the higher wages of skilled individuals (Rauch, 1993; Glaeser *et al.*, 1995; Florida *et al.*, 2012).

A solution to address this issue is to add an additional dimension so that $\mathbf{C} = (C_{cap})_{N_c \times N_a \times N_p}$. The new dimension indexed by p would count how many "skill units" of skill a are being used by city c to produce product p. This specification of the model would be more realistic and more easily interpretable. These units would be proportional to a labor force, and thus the population of the whole city would naturally be proportional to the total number of units across all skills and all products.

$$pop_c \propto \sum_{a=1}^{N_a} \sum_{p=1}^{N_p} C_{cap}. \tag{5.6}$$

However, adding this dimension not only makes the simulation of the model computationally expensive, but it also introduces the need to define rules for how skills are allocated to produce different products, e.g., through optimization, etc.. We eschew these considerations by simply assuming that the value of C_{ca} already accounts for its availability.

Another conspicuous limitation is that the model does not allow open-ended diversification and innovation since we assume that the number of skills N_a is finite and countable (see Bettencourt *et al.*, 2014 for an empirical study about this topic). For this reason, the space of possible products is also finite. This limitation, however, is not very restrictive since the number of possible products scales exponentially like 2^{N_a}. Hence, for a large number of skills, the space of products will be huge and most of it will not be explored anyway. From theoretical reasons, there is an expectation that a relation between technological innovations and the process of biological evolution, which is open-ended, exists (Arthur and Polak, 2006; Beinhocker, 2006; Thurner *et al.*, 2010; Klimek *et al.*, 2010, 2012). Consequently, a correct theory of how technological innovation occurs must clarify the commonalities and differences between the process of biological evolution. In the case of our model, its lack of open-endedness goes precisely against the expectation that both innovation and biological evolution are open-ended.

And lastly, we do not model the loss of skills. It is indeed of special interest to understand in general the migration patterns of skills in cities, for example, by measuring by the flows of people in and out of job occupations. Understanding these flows, and analyzing their impact in the production processes in cities has been an important subject of research in urban economics. Here in our study, however, we only aim to study if and how a multiplicative process can emerge from a model of production.

5.3 Simulation

For the following sections, we will use the total number of capabilities in a city as a measure of its size. As discussed in the previous section, this may or may not be linearly related to the actual population size, but is still a measure of size. Hence,

$$size_c = \sum_{a=1}^{N_a} C_{ca}. \tag{5.7}$$

5.3.1 The Binomial Model

Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011) proposed a model, the "binomial model", that we shall argue is a building block of a more general model. More precisely, we will show that their model can be viewed as a one-step iteration of a model that can be iterated several times.

The binomial model consists of filling the matrices $\mathbf{C}$ and $\mathbf{P}$ such that each element is 1 with a probability r and q, respectively, and 0 otherwise. Then it uses the Leontief production function to calculate $\mathbf{X} = \mathbf{C} \odot \mathbf{P}$.

Since the matrices $\mathbf{C}$ and $\mathbf{P}$ are binary, the matrix $\mathbf{X}$, which is denote as M_{cp} in the original papers, is also binary. The interpretation of M_{cp} is of an adjacency matrix describing a bipartite network of which products are exported by which countries. The *diversification* of a country is defined in the binomial model as the number of products the country exports:

$$k_{c,0} = \sum_{p} M_{cp}. \tag{5.8}$$

Hausmann and Hidalgo (2011) find that this measure is well fitted by lognormal and weibull distributions.

The binomial model is successful at reproducing some stylized facts of the network connecting countries and the products they export (see Hidalgo and Hausmann, 2009; Hausmann and Hidalgo, 2011). But in the context of our question (why do we see lognormals in the distributions of urban output conditional on population size?) this model can be viewed as a "bernoulli step" in a more general dynamical model. Hence, these matrices and the results they generate have a different interpretation for us.

The fact that the diversification, as defined in equation (5.8) by this simple model, is lognormally (or weibul) distributed invites a dynamic generalization of the model and the analysis of the robustness of the distributions generated. Let $\mathbf{C^{t=0}}$ be an initial matrix of zeros, and $\mathbf{C^{t=1}}$ the matrix filled as in the bernoulli step of Higalgo and Hausmann's model. Now, the next time step will consist of generating again the bernoulli step with the same probability r and adding the result to $\mathbf{C^{t=1}}$ to create $\mathbf{C^{t=2}}$. This process is repeated at each time step. It is then trivial to see that after T time steps, each element $C_{ca}^{t=T}$ is a random variable binomially distributed $\mathscr{B}(T,r)$. Note, then, that Hausmann and Hidalgo (2011)'s diversification is for us the urban output $y_c^{t=1}$ in the first iteration of the model. The question is whether $y_c^{t=T}$ is lognormally distributed, and under what conditions.

5.3.2 The Poisson Limit and Model Parameters

The physical interpretation of r in the dynamic version of the binomial model is that it represents the probability of acquiring a unit of capability over a single time-step. This

Table 5.1: Poisson Model as the Limit of the Binomial Model. Parameters of the Model in the Limit of $T \to \infty$ with $rT = \lambda$.

Parameter	**Meaning**
$N_c \in \mathbb{N}_{>0}$	Number of cities
$N_p \in \mathbb{N}_{>0}$	Number of products
$N_a \in \mathbb{N}_{>0}$	Number of skills/capabilities
$q \in (0,1)$	Bernoulli probability for filling **P**
$F_{\odot} \in \{\text{Leontief}, \text{Cobb-Douglas}, \text{Linear}\}$	Production function
$C_{ca} \sim$ i.i.d $Poisson(\lambda)$ with $\lambda \in \mathbb{R}_{>0}$	Rate of acquisition of skills for **C**

time-step could be measured in days, months or years, depending on the process, and T could represent longer periods of time. In a city, however, the acquisition of a single skill can be assumed to happen in the order of seconds, and with very small probabilities r. But one is typically interested in the accumulated statistics over periods of time much longer than seconds (e.g., a year), which means that T will be very large. To account for this, the binomial distribution which we use to simulate the acquisition of skills has a simple approximation. As $r \to 0$ and $T \to \infty$, the distribution becomes approximately Poisson. This simplifies the model and is useful for simulating different instances to explore the parameter space.

Given

$$\lambda \equiv \lim_{r \to 0, T \to \infty} Tr,$$

then $C_{ca}^{T \to \infty} \sim Poisson(\lambda)$. In this Poisson limit, λ represents the *rate* over a period of time (e.g., a year) at which skills are acquired[2]. Table 5.1 lists the parameters of the model in this limit and their meaning.

In general, the number of cities N_c in urban systems is on the order of thousands, and therefore we will fix it as $N_c = 1000$ for all following simulations.

[2]Note that there is no conflict in the units of λ, since T is the parameter of the Binomial distribution and is a number with no dimensions. In this way, λ and r share the same dimensions of a rate.

The number of products, in contrast, depends on the phenomenon we are studying and the resolution and scale at which we are looking. In a criminological context, for example, different criminal know-hows produce different types of crimes. Each type of crime is a different "product" of a different set of criminal activities within cities. Clearly, the number of different products in this context depends on our classification of crimes. We may refer to two: violent and non-violent, for example. But there may be tens of thousands or perhaps more. In the case of exports by countries, for example, Hausmann and Hidalgo (2011) chose two classification schemes, from the SITC4 dataset and the HS6 dataset (see Hidalgo and Hausmann, 2009 for the sources of the data and the details), for which $N_p = 775$ and $N_p = 5109$, respectively. Here, we will fix the number of products to $N_p = 2000$ since it seams reasonable to think that the number of products is larger than the number of cities, but we keep it on the same order of magnitude.

Lastly, we need to define the number of skills N_a. This number has the same issues as the number of products, namely, the dependence on the resolution and classification scheme. According to the way we decided to define the products that make up the matrix $\mathbf{P}$, there are $2^{N_a} - 1$ total number of possible products[3]. This number increases exponentially fast as we increase the number of skills. In principle, from the point of view of the mathematical consistency of the model, the only constraint we have is that N_p has to be smaller than the total possible number allowed by specifying N_a, i.e., $N_a \geq \left\lceil \frac{\ln(N_p+1)}{\ln 2} \right\rceil$. We will fix $N_a = 100$ to see how the model behaves with the number of skills in the order of the hundreds, even though in this specification the percentage of products actually defined out of the total possible, $N_p/(2^{N_a} - 1)$, is very small.

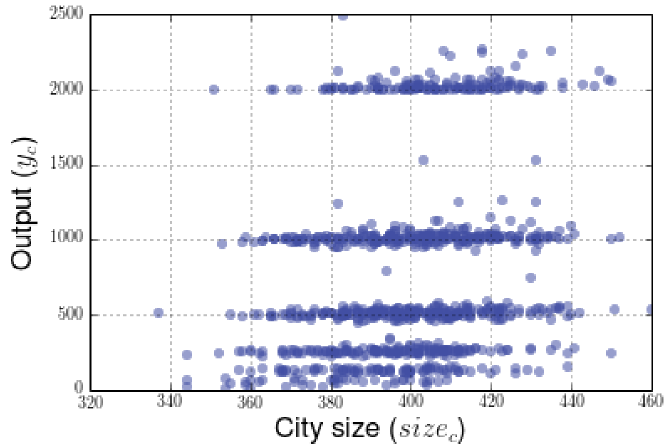


Figure 5.1: Typical Scatter Plot of Output Per City, y_c, Versus the Measure of City Size, $size_c$, for $q = 0.5$ and $\lambda = 4.0$. Here, $N_c = 1000, N_p = 2000, N_a = 100$, and $C_{ca} \sim Poisson(\lambda)$. Each Point is a City c. Interestingly, Discrete Levels of Output, Represented by "Clouds" of Points at Different Heights, Emerge from the Model Spontaneously. These Differences Are Analyzed in the Text, and Shown to Arise from a Multiplicative Effect ω. In this Figure, Levels Go Approximately from $100 \rightarrow 250 \rightarrow 500 \rightarrow 1000 \rightarrow 2000$, Which Suggests that $\omega \approx 2$.

5.4 Results

Figure 5.1 shows the relationship of the output, y_c, versus the measure of size, $size_c$, for $\{N_c = 1000, N_p = 2000, N_a = 100\}$, when $q = 0.5$ and $\lambda = 4.0$. This plot shows an important feature of the model related to the different "levels" in which the dots appear to cluster. Hence, there is a broad discretization of output at different levels that emerges from the model. Moreover, cities cluster in levels of output that increase *multiplicatively*, and we refer to this effect as ω. This is manifested as bigger and bigger jumps from one level to the next (see caption in fig. 5.1).

Note that these levels are not explained by differences in size, or more precisely, differences in the total number of capabilities. These differences arise from the differences in arrangements that the skills within cities form to produce different products. We regard this

[3] The "-1" in the expression comes from not counting the product defined as an array of zeros.

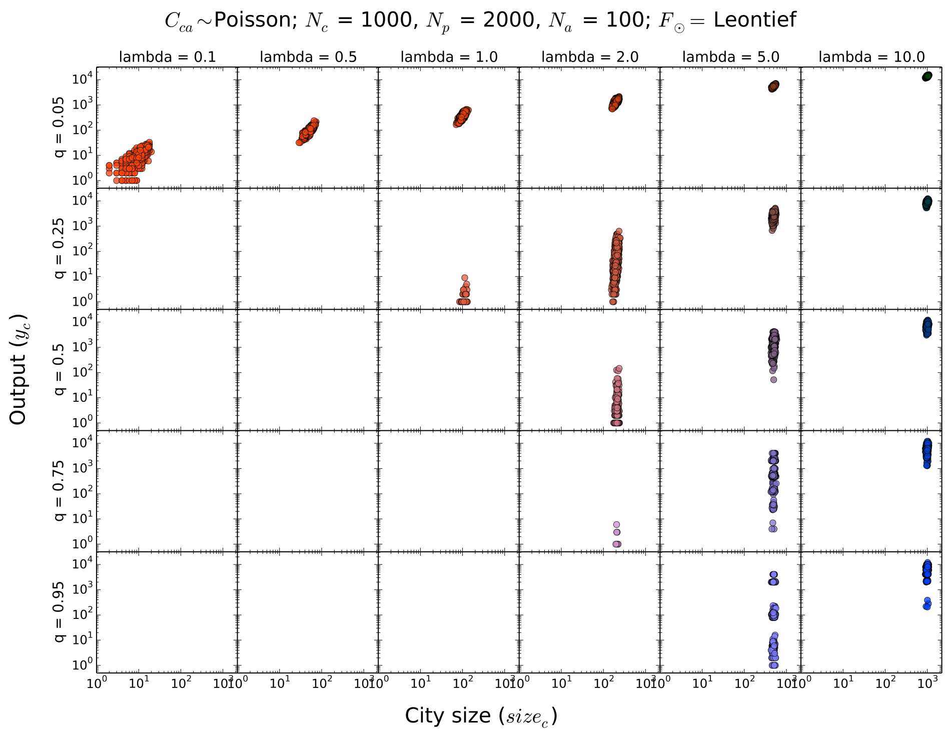


Figure 5.2: Scatter Plots of y_c, in Logarithmic Scale, Versus $size_c$, in Logarithmic Scale, for Different Combinations of the Parameters q and λ. Here, $N_c = 1000, N_p = 2000, N_a = 100$.

multiplicative feature of the simulated output, as a stepping-stone towards understanding the emergence of lognormally distributed aggregate urban output, given size.

Figure 5.2 shows an overview of different combinations of q and λ, in which the axes of all panels are of the same scales. This makes us lose the ability to see the internal structure and more particular characteristics of the scatter plots, but it provides us with a quick way of looking at the different profiles of output versus size.

The empty plots in fig. 5.2 indicate that cities were not able to produce the products. This will be explained in section 5.4.1 in more detail, but intuitively this comes from the fact that when q is large, the products become too complex to produce (i.e., too many skills required), especially if λ is small and cities have very few skills.

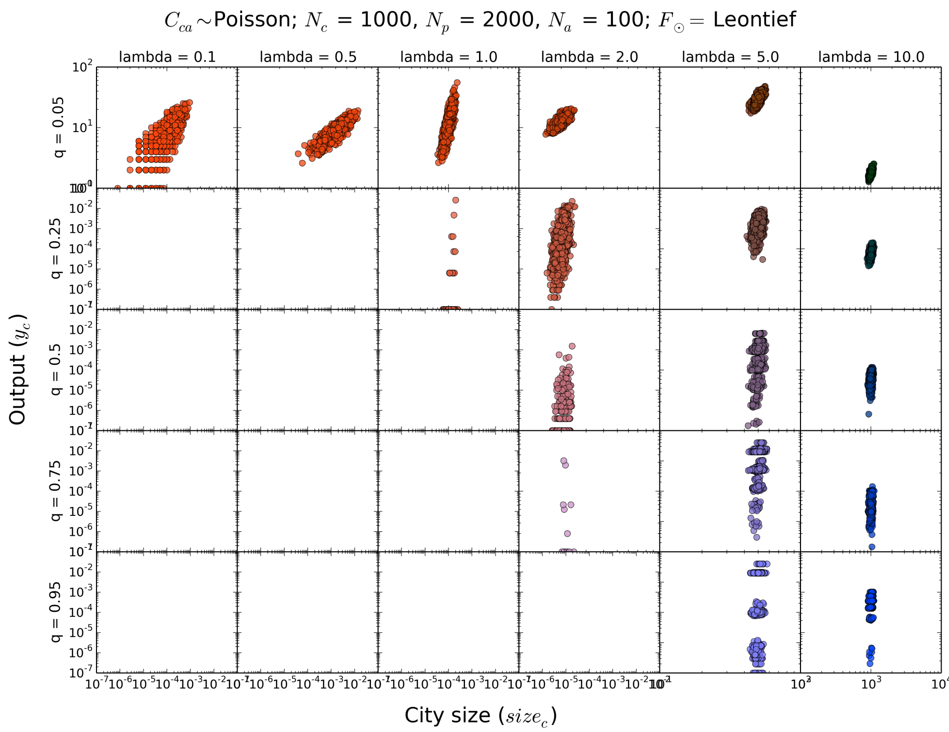


Figure 5.3: Scatter Plots of y_c, Logarithmic Scale, Versus $size_c$, in Logarithmic Scale, for Different Combinations of the Parameters q and λ. Here, $N_c = 1000, N_p = 2000, N_a = 100$.

In fig. 5.3 we have plotted the same model results, but now the subplots use their own appropriate scales for the plot axes. In this visualization of the data, the distribution of the points are more clearly visible.

Figure 5.4 plots the corresponding histograms of the output conditional on city size. We use the same methodology as in chapter 4, in which we standardize the logarithm of output for *all* size bins. Here, however the bins are linear and not logarithmic because the range of different sizes is not broad enough. Once standardized, the distribution of the values are collapsed into one single histogram. The estimates of the parameters μ and σ of the lognormal distributions for each bin are shown in appendix F.

In each histogram we show an inset with a Q-Q plot comparing the sample quantiles in the vertical axis with the theoretical quantiles in the horizontal axis of a normal distri-

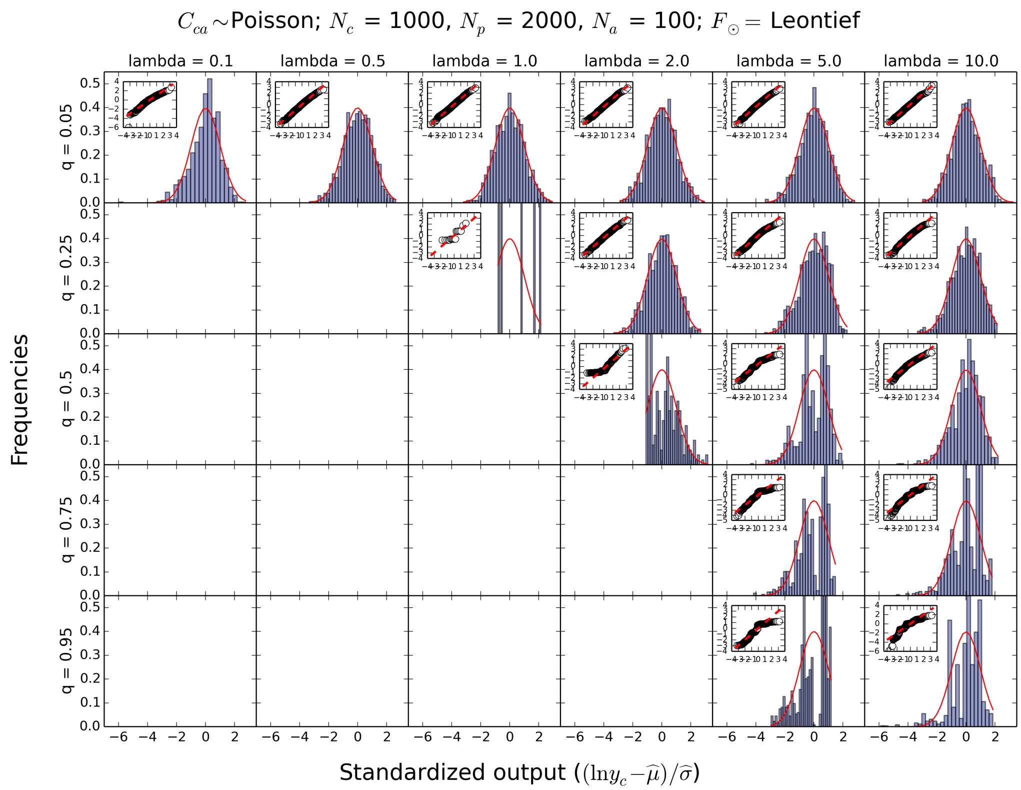


Figure 5.4: Histograms of Standardized $\ln y_c$ Conditional on $size_c$ (Binned Linearly) for Different Combinations of the Parameters q and λ. The Inset Plots Are Q-Q Plots for the Normal Distribution. If the Output is Lognormally Distributed the Dots Should Line Up with the Red Dashed Line. Here, $N_c = 1000, N_p = 2000, N_a = 100$.

bution. When the dots line up with the dashed red straight line it indicates that the distribution of $\ln(y_c)$, conditional on $size_c$, is approximately normal (i.e., that y_c is lognormally distributed). The plots in fig. 5.4 show that the distribution of $\ln(y_c)$ is more normal, i.e., y_c is more lognormally distributed, for small values of q. For large values of q the dots in the insets show a concave shape, or inverted U, indicating that the distribution of $\ln(y_c)$ has thinner right tails than a normal distribution, but heavier left tails. This is expected if y_c is starting to become more normally distributed.

From the histograms, it can also be noted that the distribution's multimodality becomes more evident as q increases above 0.5, reflecting the levels of output that are visible in fig. 5.1. This multimodality is present for all values of q and λ, and in section 5.4.1 we derive an analytical expression that explains this behavior.

In the next section we will provide with a mathematical analysis of the balance between the complexity of the products and the minimum complexity of cities required to produce those products. Analytical calculations of the multiplicative and distributional aspects of the model become rapidly intractable, but we will provide with some intuitions behind the results presented so far.

5.4.1 Analytical Calculations

The following analytical calculations are limited to the case of a Leontief production function.

Product Complexity VS. City Complexity

Parameters q and λ respectively determine the level of complexity of the products and of cities. As q increases, for example, the average number of skills that products require increases as $N_a q$. Similarly, as λ increases, the average number of skills in cities increases as $N_a \lambda$. But if products are too complex to produce (i.e., $q \approx 1$) and cities are too simple (i.e., $\lambda \approx 0$), then cities will not be able to produce any product. A question, thus, is how many skills are required to be able to produce something *at all*. In other words, *given* a level of complexity λ in a city c, what is the maximum complexity q of products such that $X_{cp} > 0$ for at least one product p among the N_p?

Mathematically, the condition to produce more than k units of production in more than n_p number of products, on average, is:

$$N_p \Pr\left\{X_{cp} > k\right\} \geq n_p. \tag{5.9}$$

The minimum condition, then, is that there is at lease one product in which a city c is able to produce one or more units:

$$N_p \Pr\left\{X_{cp} > 0\right\} \geq 1. \tag{5.10}$$

For this, we need to derive the distribution of X_{cp}.

Assume city c has the following skills: $\vec{c} = (C_{c1}, C_{c2}, \ldots, C_{cN_a})$. Given a product p requiring M skills to produce it, $a_1, \ldots, a_M \in \{1, 2, \ldots, N_a\}$, then X_{cp} can be written as the minimum of M independent and identically distributed random variables. In our model $X_{cp} = \min\{C_{ca_1}, \ldots, C_{ca_M}\}$, where $C_{ca_l} \sim Poisson(\lambda)$, for all $l = 1, \ldots, M$. From the way we decided to generate the matrix $\mathbf{P}$, the variable M is a binomial random variable, distributed $M \sim \mathscr{B}(N_a, q)$. Hence,

$$\Pr\left\{X_{cp} \leq k\right\} = \sum_{l=0}^{N_a} G_{k|l} \, \Pr\{M = l\}, \tag{5.11}$$

where $G_{k|l}$ is the distribution $\Pr\left\{X_{cp} \leq k \mid \text{product } p \text{ with } M = l\right\}$. This distribution is in turn given by

$$G_{k|l} = 1 - \Pr\left\{X_{cp} > k \mid \text{product } p \text{ with } M = l\right\}. \tag{5.12}$$

That $X_{cp} > k$ in equation (5.12) means that all the $M = l$ values $C_{ca_1}, \ldots, C_{ca_l}$ have to be larger than k. Since they are mutually independent and identically distributed, then the second term at the right-hand side of equation (5.12) is just the product of the distribution of C_{ca}:

$$\begin{aligned} G_{k|l} &= 1 - \left[\Pr\{C_{ca_1} > k\}\right]^l \\ &= 1 - \left[1 - \Pr\{C_{ca_1} \leq k\}\right]^l \\ &= 1 - \left[1 - R_k\right]^l, \end{aligned} \tag{5.13}$$

where R_k denotes the cumulative distribution of the elements C_{ca}. For our poisson model, R_k is the distribution of a poisson random variable (see table 5.1), and $R_k = Q(\lfloor k+1 \rfloor, \lambda)$, where Q is the regularized gamma function defined as

$$Q(s,\lambda) = \frac{\Gamma(s,\lambda)}{\Gamma(s,0)}, \qquad \text{where } \Gamma(s,x) = \int_x^\infty t^{s-1}\mathrm{e}^{-t}\mathrm{d}t. \tag{5.14}$$

Putting together equation (5.11) and equation (5.13), we get

$$\Pr\left\{X_{cp} \le k\right\} = \sum_{l=0}^{N_a} \left(1-[1-R_k]^l\right)\ \Pr\{M=l\}, \tag{5.15}$$

Equation (5.15) simplifies into

$$\begin{aligned}\Pr\left\{X_{cp} \le k\right\} &= 1-\sum_{l=0}^{N_a}[1-R_k]^l\ \Pr\{M=l\} \\ &= 1-\sum_{l=0}^{N_a}[1-R_k]^l\ \binom{N_a}{l} q^l(1-q)^{N_a-l} \\ &= 1-[q(1-R_k)+(1-q)]^{N_a}, \end{aligned} \tag{5.16}$$

where we have used the binomial expansion in the last step. Finally, the exact expression for the distribution of X_{cp} is:

$$\Pr\left\{X_{cp} \le k\right\} = 1-(1-qR_k)^{N_a}. \tag{5.17}$$

Going back to the question about the balance between the complexity of skills in cities and the complexity of products, we can now use equation (5.17) in equation (5.10), which becomes $N_p(1-q\mathrm{e}^{-\lambda})^{N_a} \ge 1$ (where we have used the fact that $R_0 = Q(1,\lambda) = \mathrm{e}^{-\lambda}$ in our model). After some rearrangements, we get the following condition:

$$q \le \left(1-N_p^{-1/N_a}\right)\mathrm{e}^{\lambda}. \tag{5.18}$$

Using $N_p = 2000$ and $N_a = 100$, fig. 5.5 plots the combinations of λ and q that are necessary for a given city to produce at least one of the products.

Equation (5.18) shows that, in this model, an increase in the number of skills of a city has an exponential effect on the complexity of the products it can produce.

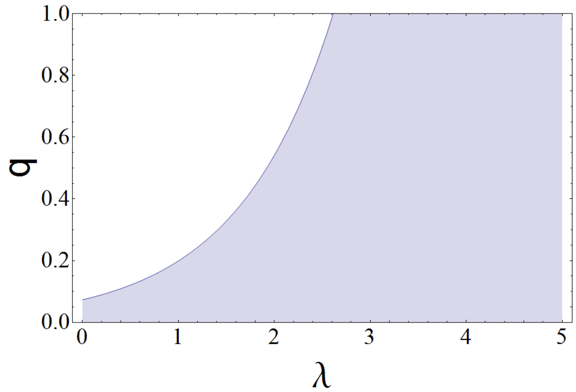


Figure 5.5: Balance of Between Complexity of Products and Cities. The Shaded Region Describes the Combinations of the Parameters q and λ for $N_p = 2000$ and $N_a = 100$ that would Allow Cities to Produce Something.

Multiplicative Effects in Aggregate Output

Knowing that X_{cp} in our model can only take values on the natural numbers, equation (5.17) can be used to calculate the expected value of production per product:

$$\begin{aligned} \mathrm{E}\left[X_{cp}\right] &= \sum_{k=0}^{\infty} \Pr\left\{X_{cp} > k\right\} \\ &= \sum_{k=0}^{\infty} \left(1 - qR_k\right)^{N_a}. \end{aligned} \tag{5.19}$$

Since the term $(1 - qR_k)$ is less than 1 and tends to 0 as $k \to \infty$, and N_a is large, the sum in equation (5.19) can be approximated by taking just the first term of the summation, to get $\mathrm{E}\left[X_{cp}\right] \approx (1 - qR_0)^{N_a}$. For our Poisson model, $R_0 = Q(1, \lambda) = \mathrm{e}^{-\lambda}$, and this approximation yields the following relation:

$$\mathrm{E}\left[X_{cp}\right] \approx \left(1 - \frac{q}{\mathrm{e}^{\lambda}}\right)^{N_a}. \tag{5.20}$$

Equation (5.20) is another instance where it can be seen that a linear increase in product complexity q can be offset with an exponential effect of an increase in the complexity of skills in the city λ.

Our model shows that producing a complex product increases the probability that a city produces other less complex products. As a consequence, the probability $\Pr\{X_{cp} > 0 \mid X_{cp'} > 0\}$ that a city produces product p, given that it already produces other product $p' \neq p$, can be positive. Because of this, we cannot use equation (5.20) to approximate the average output y_c as the independent sum of the average production of each product. That is to say, $\mathrm{E}[y_c|\text{a particular city}] \neq N_p \mathrm{E}[X_{cp}]$.

To address this problem, let $C_{(1)} \leq C_{(2)} \leq \ldots \leq C_{(N_a)}$ denote the ordered values of the elements of a city c, $\vec{c} = (C_{(1)}, C_{(2)}, \ldots, C_{(N_a)})$. The values $C_{(1)}, \ldots, C_{(N_a)}$ are called the *order statistics* of the original i.i.d. random variables $C_1, \ldots, C_{N_a}$ (throughout this section we will drop the index c of the elements C_{ca} and keep only the index a of the skills). Let $\vec{p} = (P_1, P_2, \ldots, P_{N_a})$ be a product. Note that since the values in $\vec{p}$ are independent of the values in $\vec{c}$, the ordering does not affect how we index the elements in $\vec{p}$.

Now, reading the vector $\vec{p}$ from left to right, let the random variable A denote the index of the *first* element in $\vec{p}$ that is a 1. For example, if $\vec{p} = (0,0,1,0,1,\ldots,0)$, then A takes the value $A = 3$. Since all elements of $\vec{p}$ are a i.i.d. bernoulli random variables with probability q, the random variable A is a geometric random variable

$$\Pr\{A = a\} = \begin{cases} (1-q)^{a-1}q, & \text{for } a = 1, \ldots, N_a \\ 0, & \text{otherwise.} \end{cases} \tag{5.21}$$

Now, we consider the skills within a city c as given, and we calculate the average output. For this, we add over all skills the average number of products whose index A is a times the value of the skill $C_{(a)}$ that c has for that index:

$$\begin{aligned}
\mathrm{E}\left[y_c \mid \vec{c}\right] &= \sum_{a=1}^{N_a} \left(N_p \Pr\{A = a\}\right) \, C_{(a)} \\
&= \sum_{a=1}^{N_a} N_p (1-q)^{a-1} q C_{(a)} \\
&= q N_p \sum_{a=1}^{N_a} C_{(a)} (1-q)^{a-1} \\
&= q N_p \left(C_{(1)} + C_{(2)}(1-q) + C_{(3)}(1-q)^2 + \ldots + C_{(N_a)}(1-q)^{N_a - 1}\right). \qquad (5.22)
\end{aligned}$$

It is from equation (5.22) that we can understand the multiplicative effect in the aggregate output that emerges from the model, and where the balance between a normal and lognormal distribution lies.

Let us denote $\mathrm{E}\left[y_c \mid \vec{c}\right] \equiv \bar{y}(\vec{c})$. Taking partial derivatives from equation (5.22), we get

$$\frac{\partial \bar{y}(\vec{c})}{\partial C_{(a)}} = q N_p (1-q)^{a-1}, \quad \text{for } a = 1, \ldots, N_a. \qquad (5.23)$$

What equation (5.23) reveals is, first, that the average effect on the output of an increase in one of the skills depends critically on *which* skill is increased. And second, since $1 - q < 1$, that the effect decays exponentially with a. Comparing the effect of one index a to the next $a+1$, yields a ratio that quantifies the multiplicative effect, which we denote with the greek letter 'omega':

$$\begin{aligned}
\omega &= \frac{\partial \bar{y}(\vec{c}) / \partial C_{(a)}}{\partial \bar{y}(\vec{c}) / \partial C_{(a+1)}} \\
&= \frac{1}{1-q}. \qquad (5.24)
\end{aligned}$$

The prediction is that the multiplicative effect ω will be greater as $q \to 1$. This is exactly what is observed in fig. 5.4: as q approaches 1, the distribution becomes multimodal, and the difference between the levels (i.e., the "modes") increases. It also explains why in fig. 5.1 the observed multiplicative effect is 2. Namely, that $\omega = 2$ for that particular plot because $q = 0.5$.

Multiplicative VS. Additive Effects

The other side of equation (5.22) is that it represents a sum of N_a terms. If we now let $\vec{c}$ to be a random vector, we expect that the distribution of y_c to approximate a normal distribution as $N_a \to \infty$, even though the terms are not identically and independently distributed (Frank and Smith, 2011), since the correlations between ordered statistics are typically weak (Arnold *et al.*, 2008) for thin-tailed distributions such as the Poisson. However, in our simulations we held fixed $N_a = 100$, and we are here more concerned with the transition from multiplicative to additive effects induced by the other parameters.

Directly deriving the distribution of y_c, however, is not trivial given that it involves accounting for the distribution of sums of the order statistics $C_{(a)}$.[4] But we can study the distributional properties of equation (5.22) under some limiting circumstances.

The condition for y_c to be a lognormal random variable is that $\ln(y_c)$ must be normally distributed. According to this, and by taking the logarithm of equation (5.22), what are the conditions for

$$\ln(q\,N_p) + \ln\left(\sum_{a=1}^{N_a} C_{(a)}(1-q)^{a-1}\right)$$

to be normally distributed?

A first approximation is to let $q \to 1$. All the terms except the first ($a = 1$) can be neglected in the sum, and we are left with

$$\ln(\bar{y}(\vec{c})) \approx \ln(qN_p) + \ln\left(C_{(1)}\right), \quad \text{for } q \to 1. \tag{5.26}$$

[4]The distribution $\Pr\{C_{(a)} \leq k\}$ can be derived as follows. The event that $C_{(a)} \leq k$ can occur if j variables among $C_1, \ldots, C_{N_a}$ are less or equal than k, *and* if the rest $N_a - j$ are larger than k. But note that there have to be at least a values less or equal than k. Therefore, this can happen for all values $j = a, a+1, \ldots, N_a$. Thus,

$$\Pr\{C_{(a)} \leq k\} = \sum_{j=a}^{N_a} \binom{N_a}{j} (R_k)^j (1-R_k)^{N_a - j}, \tag{5.25}$$

where $R_k = \Pr\{C_a \leq k\}$.

If we suppose that the elements of $\vec{c}$ have a distribution such that the minimum, $C_{(1)}$, is lognormally distributed, then $\bar{y}(C_{(1)})$ will be lognormally distributed.

5.5 Discussion

To summarize, the motivations of this project were that: (1) we see economic output approximately lognormally distributed that suggests a multiplicative process, and (2) that knowledge and skills are known to have an important role in the productivity of regions. Hence, we investigated a model of skills in cities from which we found a multiplicative effected emerged. We believe this multiplicative effect is the first step to understand lognormal distributions.

The model we proposed and analyzed, is a generalization of the model proposed by Hidalgo and Hausmann (2009) and Hausmann and Hidalgo (2011). It is a discrete model that in which the aggregate total output that cities produced is clustered in different levels, increasingly separated by a multiplicative effect. The conclusion that a multiplicative effect necessarily implies a lognormal distribution is not warranted. Instead, identifying a multiplicative effect is the first step of many, for understanding the emergence of lognormals. Part of this is answered by our computational results, which show under which conditions the distributions of output conditional on city size resembles more closely a lognormal. Therefore, our results can be summarized not so much as whether we were able to reproduce lognormality, but rather, that we were able to identify a multiplicative effect. A question for the future is then, what are the other elements that must be in place that generate lognormal outputs.

We focused our attention to the analysis of the parameters q and λ. Parameter q gave us a measure of the complexity of the products, since the average number of skills required to produce a product is given by $N_a q$. In the same way, parameter λ gave us a measure of the complexity of cities, since the average number of skills in a city is given by $N_a \lambda$. These notions of product and city *complexity* are limited by the simplicity of the model, and they may not reflect the real complexity of cities and products. Nevertheless, they provide a useful language to think about the underlying causes of the differences in economic performance that emerge between cities of the same population size. Namely, the differences in the complexity of their internal economy *and* the complexity of the products that they each produce.

We showed through figs. 5.2 and 5.3 that there are two balances at play. On the one hand, there is a balance between the complexity of the products and the complexity of cities. And on the other, there is a balance between normal and lognormal distributions. Or in other words, between additive and multiplicative processes.

We expressed mathematically the balance between the complexities of products and cities by equation (5.18). This equation stated that the maximum value of q that allows a city with parameter λ to produce something at all increases exponentially with λ. The condition to produce something can be written more generally as

$$N_p\,(1-qR_k)^{N_a} \geq n_p, \tag{5.27}$$

which states the condition in which a city produces on average more than n_p products with an intensity X_{cp} greater than k. Equation (5.18) is a particular case for which $n_p = 1$ and $k = 0$. With these values $R_0 = \mathrm{e}^{-\lambda}$. This exponential term that appears in equation (5.18) and also in equation (5.20) comes from the exponential in the poisson probability mass function. It is thus possible that similar results may or may not hold for other probability distributions, and should be the subject of future work.[5]

[5] A place to start are the exponential family distributions (Marin and Robert, 2007) for which the probability function can be written as

$$f_{C_{ca}}(k \mid \vec{\theta}) = h(k)\mathrm{e}^{\vec{\theta}\cdot\vec{R}(k)-\Psi(\vec{\theta})},$$

where $\vec{\theta}$ is the vector of parameters.

One could generalize equation (5.18) to hypothesize that

$$CP_{\max} = CP_0 \, e^{CC/CC_0}, \tag{5.28}$$

where the $CP_{\max}$ stands for a measure of the maximum complexity of a product that a city with a complexity of CC would be able to produce. The terms CP_0 and CC_0 are just scale coefficients. In our particular model, for instance, $CC = N_a \lambda$, $CP_0 = N_a \left(1 - N_p^{-1/N_a}\right)$, and $CC_0 = N_a$.

Although equations like (5.20) or (5.28) define relationships between average quantities, they provide a hint for possible multiplicative mechanisms behind the lognormal distributions of output that we have observed in urban areas of comparable population sizes.

This multiplicative effect is most evident, for example, in fig. 5.1, in which several discrete levels of production can be seen, separated from each other by a multiplicative factor (which in case of fig. 5.1 seems to be 2). But this multiplicative effect is in tension with the additive aspect that implies an aggregate measure of total output. To understand all this we derived an expression of the average output for a given city and its skills (equation (5.22)). The balance between multiplicative and additive processes, as is seen in fig. 5.4, is that as products become more complex, the multiplicative effects are larger. At the same time, however, this multiplicative effect discretizes the distribution, breaking it into multiple modes. Again, the analytical expression of such behavior in the model is provided by equation (5.22).

We quantified the multiplicative effect by ω. This term goes like $\frac{1}{1-q}$. In this sense, equation (5.22) supports the intuitive idea that as products in urban economies require more and more skills to be produced and become more complex, small changes in some of the available skills to the city can have a cascading effect on the aggregate economic output.

If the products require few skills to be produced, then $\omega \to 1^+$, and total output becomes the sum of many factors which contribute about equally to the sum. This situation leads to a normal distribution of output.

5.6 Concluding Remarks

We have proposed a model, the Poisson model, to study the statistical patterns of economic output at the subnational level. In particular, our goal was to explain the *lognormal-like* distributions that urban areas of comparable population size display across different measures of output, such as creative and inventive activities (see chapter 4), homicides (see chapter 3), and others (see Bettencourt *et al.*, 2010; Alves *et al.*, 2013a,b, 2014; Mantovani *et al.*, 2013).

Our model was an extension to what Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011) proposed to explain the patterns of national exports. Our aim and interpretations of the model, however, were different from Hidalgo and Hausmann's. While the original binomial model of Hidalgo and Hausmann was static in time, ours has a time dimension implicit to it. But more importantly, we modeled not only *what* cities produce, but *how much*, and we studied the hypothesis that differences in skills between urban economies affect multiplicatively the aggregate urban output.

The model is grounded on the idea that the total output of a city is the aggregate production across many products, each of which require a varying set and number of capabilities to produce them. In this way, if a city has few skills, it might not have the required capabilities and skills to produce many of the products. But on the same token, acquiring a single skill, or even replacing one for another, can have an cascading effect in the number of products that a city may be able to produce, and as a consequence, a multiplicative effect in its total output. In a way, our results formalize a statement by Florida (2011) that we quoted in chapter 1: "[w]hen talented and creative people come together, the multiplying effect is exponential; the end result is much more than the sum of the parts. Clustering makes each of us more productive–and our collective creativity and economic wealth grow accordingly" (Florida, 2011, p. 193).

Naturally, our model is a simplification of the production processes in cities. Nonetheless, the model has helped us to understand some aspects of the *distributional patterns* that describe differences in urban socio-economic output. This is one aspect of cities that has recently received little attention. Our study is the first to provide a mechanism that may potentially explain these distributional characteristics. Not surprisingly, the model also opens up new questions for further investigation. For example, questions about the apparent open-endedness of the processes of innovation, or about the generation of value not only at the level of the whole city, but at the level of skills and individuals (i.e., who creates wealth? who acquires the value?). Also, we simplified how products were defined, and a question for future work is to investigate the additional sources of heterogeneity that may arise from generalizing the matrix of products $\mathbf{P}$ to have other values other than 0's and 1's.

The model also offers some predictions. One is that since the multiplicative effects will become stronger as cities become more complex and diverse, and as the products they produce also become more complex by requiring larger sets of skills and capabilities, one may see then a divergence between cities in terms of their economic output. And another is that cities with economies with low diversity of skills (i.e., low complexity in their products) would not display the multiplicative effects, and therefore, would not manifest lognormal distributions in their output. This is consistent with the intuitions from Florida (1995) that we quoted in the introductory section. Thus, we hypothesize that urban systems whose economies strongly relies on the exploitation of natural resources, for example, would not display lognormal variations in urban output.

It is critical that the robustness of these assumptions is assessed in future work, and that the predictions of our model are contrasted against empirical data. Only in this way we can make progress. This chapter has advanced our understanding of how skill specialization at the level of individuals and skill diversification at the level of urban areas generate economic value. As people rapidly concentrate in urban areas, cities face the urgent problem of accommodating the diversity of individuals in which they can successfully participate in the large array of overlapping networks that constitutes the cities (Jacobs, 1969; Bettencourt, 2013; Hausmann, 2013). In this respect, we believe it is essential to understand how cities integrate human skills and knowledge, and how this is reflected in their aggregate productivity. Understanding this problem could have a particularly positive impact in developing countries, where urbanization occurs rapidly while education, knowledge, and skills remain low (The World Bank, 2013).

Chapter 6

SUMMARY AND CONCLUSION

Urban dynamics, as can be seen, are the ultimate noisy social science problem. (Storper *et al.*, 2012, p. 4)

In this dissertation I have taken the view that strong fluctuations, broad variability, and pervasive heterogeneity seem to be at the heart of urban systems, within and across cities. This is not to say that cities are objects of complete randomness. On the contrary, we take the view that the noisiness of urban life gives rise to structured and systematic regularities, although not in the form of exact relations, but rather in the form of statistical relations and probabilistic statements. Accordingly, I have argued that a deeper understanding of the mechanisms and generative processes in cities can be gained by addressing the full statistical characterization of urban quantities. I have used quantities of interest in cities, such as crime, invention, and creativity to support this.

In chapter 2 we presented a parsimonious null model of an urban system to show that both the broad variance in urban productivity across cities and its average allometric scaling with population size can be have a different interpretation than the one implied by the conventional econometric models. Although we have not given economic micro-foundations to bolster the assumptions of the model, it contributes to the establishment of urban scaling theory in two ways: **(1)** it formalizes the use of population size in a statistical account of urban scaling. And **(2)**, it formally shows how the internal heterogeneities in cities determine the statistical properties of the aggregate output at a systemic level. The model shows that when the generative process of individual wealth creation within cities creates extreme differences in individual productivity, then automatically larger places will be on the aggregate more productive than smaller places, even though intrinsically they are not. The reason behind this result is a violation of the Law of Large Numbers. We use this result to formalize the question about when to use *per-capita* measures of urban output. The decision to use either of these measures depends on the statistical distribution describing the individual variables from which per-capita and aggregate measures are constructed. We derived conditions, holding some reasonable assumptions about the underlying distributions, about the size of socio-economic aggregations of people above which per-capita measures can have meaningful interpretations.

The take-home message of the null model is that given the finiteness of cities and their known internal heterogeneities, one must be aware that violations of the Law of Large Numbers have an important effect on our statistics at the aggregate level. When this violation occurs, the whole is typically more than the sum of the expected mean of the parts.

We then studied the variability of homicides, a highly stochastic urban metric, to understand the statistical relationship between two broad regularities: urban scaling and Zipf's law. This relationship is expressed by a co-dependence between the urban scaling exponents and the exponents of the Pareto laws describing the marginal distributions of population size and homicides. This relationship between urban scaling and the marginal distributions was mediated by the distribution of homicides conditional on population size. This conditional distribution was found to be well fitted by a lognormal distribution. The fact that we found lognormal variations around the urban scaling law suggested a stochastic multiplicative process generating such fluctuations.

Then, we used this methodological framework to quantify the statistical constraints on the creative and inventive activities in urban areas. This study is important in the context of public policy and decision making, in which we need to understand how probable or improbable is to increase the skilled workforce of a city in order to enhance its economic productivity.

Finally, we addressed the question of how variability in the aggregate productivity of cities, conditional on population size, results from the variability in the diversity of skills that cities possess. And the particular question was if, and how, this variability emerges from a multiplicative process. The expectation was that urban areas not only need a diversity of skills to enhance productivity, but also the "right" mix of skills. And in this way, having the right ingredients to produce something results in a multiplicative process. Following the logic of this argument we explored and expanded the recent theoretical framework proposed by Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011). Although the model was originally designed to explain the exports of countries in terms of their diversification, we used the conceptual framework presented in Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011) to model how multiplicative effects can emerge in urban production processes. And we found, through computational simulations and some analytical calculations, that the increases in skill endowments of cities result in an exponential increase in the number of products that cities can produce. And we also found that changes in the portfolio of skills result in multiplicative changes in total output.

Since the seminal paper by Glaeser *et al.* (1992), important results have revealed an array of statistical associations between presumed input factors and economic outputs. But to understand the obstacles to economic development, we need to understand better the role of *size*, *heterogeneity* and *structure* in cities, which are difficult to treat using conventional regression analysis. The work presented here contributes a small grain to change this paradigm. We have presented some results to argue that these three notions require an approach whose objects of study are the probabilistic distributions, and stochastic dynamics, of urban matters. This is also important for public policies, which we also argue need to incorporate in their discourse a distributional perspective. In part because it matters not only what the averages and variances of what we measure are, but also their overall constraints characterized through their distributions.

REFERENCES

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

Agrawal, A. K., "Innovation and the Growth of Cities (Book review)", Journal of Economic Geography **3**, 4, 458–461, URL `http://search.ebscohost.com.ezproxy1.lib.asu.edu/login.aspx?direct=true&db=eoh&AN=EP11520882&site=ehost-live` (2003).

Aiken, L. S. and S. G. West, *Multiple regression: Testing and interpreting interactions* (Sage, Newbury Park, California, 1991).

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

Alves, L., H. Ribeiro, E. Lenzi and R. Mendes, "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, URL `http://www.sciencedirect.com/science/article/pii/S0378437114003641` (2014).

Alves, L. G., H. V. Ribeiro and R. S. Mendes, "Scaling laws in the dynamics of crime growth rate", Physica A: Statistical Mechanics and its Applications **392**, 11, 2672 – 2679, URL `http://www.sciencedirect.com/science/article/pii/S0378437113001416` (2013a).

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

Andriani, P. and B. McKelvey, "Beyond gaussian averages: redirecting international business and management research toward extreme events and power laws", Journal of International Business Studies **38**, 7, 1212–1230 (2007).

Anscombe, F. J., "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions", Biometrika **37**, 3/4, 358–282, URL `http://links.jstor.org/sici?sici=0006-3444%28195012%2937%3A3%2F4%3C358%3ASTOTNB%3E2.0.CO%3B2-L` (1950).

Arcaute, E., E. Hatna, P. Ferguson, H. Youn, A. Johansson and M. Batty, "City boundaries and the universality of scaling laws", arXiv preprint arXiv:1301.1674 (2013).

Arnold, B. C., N. Balakrishnan and H. N. Nagaraja, *A First Course in Order Statistics* (Society for Industrial and Applied Mathematics, 2008), URL `http://epubs.siam.org/doi/abs/10.1137/1.9780898719062`.

Arthur, W. B. and W. Polak, "The evolution of technology within a simple computer model", Complexity **11**, 1117–1119 (2006).

Bacon, E. N., *Design of Cities: Revised Edition* (Penguin (Non-Classics), New York, NY, 1976).

Baek, S. K., S. Bernhardsson and P. Minnhagen, "Zipf's law unzipped", New J. of Phys. **13**, 043004, URL `http://iopscience.iop.org/1367-2630/13/4/043004` (2011).

Bahar, D., R. Hausmann and C. A. Hidalgo, "Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?", Journal of International Economics **92**, 111–123 (2014).

Banerjee, A. and V. M. Yakovenko, "Universal patterns of inequality", New Journal of Physics **12**, 075032 (2010).

Batty, M., *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals* (The MIT Press, London, 2005).

Batty, M., "Rank clocks", Nature **444**, 7119, 592–596 (2006).

Batty, M., "The Size, Scale, and Shape of Cities", Science **319**, 769 (2008).

Batty, M., "Building a science of cities", Cities **29**, S9–S16 (2012).

Batty, M. and P. A. Longley, *Fractal Cities: A Geometry of Form and Function* (Academic Press, San Diego, 1994).

Beaudry, C. and A. Schiffauerova, "Who's right, Marshall or Jacobs? the localization versus urbanization debate", Research Policy **38**, 318–337 (2009).

Beaulieu, N. C., A. A. Abu-Dayya and P. J. McLane, "Estimating the distribution of a sum of independent lognormal random variables", Communications, IEEE Transactions on **43**, 12, 2869 (1995).

Beinhocker, E. D., *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics* (Harvard Business School Press, 60 Harvard Way, Boston, MA, 2006).

Berk, R., "What you can and cant properly do with regression", Journal of Quantitative Criminology **26**, 4, 481–487 (2010).

Berry, B. J. L., "Cities as systems within systems of cities", Papers in Regional Science **13**, 1, 147–163 (1964).

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

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

Berube, A., "All Cities Are Not Created Unequal", Tech. Rep. 52, Brookings, metropolitan Opportunity Series (2014).

Bettencourt, L. M., J. Lobo and G. B. West, "The Self Similarity of Human Social Organization and Dynamics in Cities", in "Complexity Perspectives in Innovation and Social Change", edited by D. L. et al., pp. 221–236 (Springer, 2009).

Bettencourt, L. M. A., "The Origins of Scaling in Cities", Science **340**, 1438 (2013).

Bettencourt, L. M. A., J. Lobo, D. Helbing, C. Kühnert and G. B. West, "Growth, innovation, scaling, and the pace of life in cities", Proc. Natl. Acad. Sci. U.S.A. **104**, 17, 7301–7306, URL `www.pnas.org/cgi/doi/10.1073/pnas.0610172104` (2007a).

Bettencourt, L. M. A., J. Lobo and D. Strumsky, "Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size", Research Policy **36**, 1, 107–120, URL `http://www.sciencedirect.com/science/article/pii/S0048733306001661` (2007b).

Bettencourt, L. M. A., J. Lobo, D. Strumsky and G. B. West, "Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities", PLoS ONE **5**, 11, e13541, URL `http://dx.doi.org/10.1371%2Fjournal.pone.0013541` (2010).

Bettencourt, L. M. A., J. Lobo and H. Youn, "The hypothesis of urban scaling: formalization, implications and challenges", ArXiv:1301.5919v1 [physics.soc-ph] 24 Jan 2013 (2013).

Bettencourt, L. M. A., H. Samaniego and H. Youn, "Professional Diversity and the Productivity of Cities", Scientific Reports **4**, 5393, URL `http://www.nature.com/srep/2014/140623/srep05393/full/srep05393.html` (2014).

Bettencourt, L. M. A. and G. B. West, "A unified theory of urban living", Nature **467**, 912–913 (2010).

Black, D. and V. Henderson, "Urban evolution in the USA", Journal of Economic Geography **3**, 4, 343–372 (2003).

Bouchaud, J.-P. and A. Georges, "Anomalous Diffusion in Disordered Media: Statistical Mechanisms, Models and Physical Applications", Physics Reports **195**, 4 & 5, 127–293, URL `http://www.sciencedirect.com/science/article/pii/037015739090099N` (1990).

Brummitt, C. D., A. Gomez-Lievano, N. Goudemand and G. Haslam, "Hunting for keys to innovation: The diversity and mixing of occupations do not explain a city's patent and economic productivity", in "Complex Systems Summer School Proceedings, Santa Fe Institute", pp. 1–13 (2012).

Calvó-Armengol, A. and Y. Zenou, "Social Networks and Crime Decisions: The Role of Social Structure in Facilitating Delinquent Behavior", International Economic Review **45**, 939–958 (2004).

Clauset, A., C. R. Shalizi and M. E. J. Newman, "Power-law distributions in empirical data", SIAM Review **51**, 4, 661–703 (2009).

Crosby, R. W., *Cities and Regions as Nonlinear Decision Systems*, no. 77 (Westview Press, 1984).

Curry, L., "The Random Spatial Economy: An Exploration in Settlement Theory", Annals of the Association of American Geographers **54**, 1, 138–146 (1964).

Deutsch, D., *The Beginning of Infinity: Explanations That Transform the World* (Penguin Group, New York, 2011).

Duranton, G. and D. Puga, "Micro-foundations of urban agglomeration economies", Handbook of regional and urban economics **4**, 2063–2117 (2004).

Embrechts, P., C. Klüppelberg and T. Mikosch, *Modelling extremal events: for insurance and finance*, vol. 33 (Springer, 1997).

Fischer, C. S., *The Urban Experience* (Harcourt, New York, NY, 1984).

Flanagan, W. G., *Urban Sociology: Images and Structure* (Rowman & Littlefield, Lanham, MD, 2010), 5th edn.

Florida, R., "Toward the learning region", Futures **27**, 5, 527–536 (1995).

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

Florida, R., *The Rise of the Creative Class-Revisited: 10th Anniversary Edition–Revised and Expanded* (Basic Books, New York, 2011).

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

Florida, R., C. Mellander, K. Stolarick and A. Ross, "Cities, skills and wages", Journal of Economic Geography **12**, 2, 355–377 (2012).

Frank, S. A., "The common patterns of nature", J. Evol. Biol. **22**, 8, 1563–1585 (2009).

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

Frankhauser, P., *La Fractalité des Structures Urbaines* (Anthropos, Economica, Paris, 1994).

Fujita, M., P. Krugman and A. J. Venables, *The Spatial Economy: Cities, Regions, and International Trade* (The MIT Press, Cambridge, MA, 2001).

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

Gabaix, X., "Power Laws in Economics and Finance", Annual Review of Economics **1**, 255–293 (2009).

Gaviria, A., “Increasing returns and the evolution of violent crime: the case of Colombia”, Journal of Development Economics **61**, 1–25 (2000).

Giesen, K., A. Zimmermann and J. Suedekum, “The size distribution across all cities - Double Pareto lognormal strikes”, Journal of Urban Economics **68**, 3, 129–137 (2010).

Glaeser, E. L., “Cities, Agglomeration, and Spatial Equilibrium”, OUP Catalogue (2008).

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

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman and A. Shleifer, “Growth in cities”, Journal of Political Economy **100**, 6, 1126–1152 (1992).

Glaeser, E. L., M. Resseger and K. Tobio, “Inequality in Cities”, Journal of Regional Science **49**, 4, 617–646 (2009).

Glaeser, E. L., B. Sacerdote and J. A. Scheinkman, “Crime and Social Interactions”, The Quaterly Journal of Economics **111**, 2, 507–548, URL `http://www.jstor.org/stable/2946686` (1996).

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

Glaeser, E. L., J. A. Scheinkman and A. Shleifer, “Economic growth in a cross-section of cities”, Journal of Monetary Economics **36**, 1, 117–143 (1995).

Goertzel, T. and T. Kahn, “The Great São Paulo Homicide Drop”, Homicide Studies **13**, 4, 398–410, URL `http://hsx.sagepub.com/content/13/4/398` (2009).

Gomez-Lievano, A., H. Youn and L. M. A. Bettencourt, “The Statistics of Urban Scaling and Their Connection to Zipf’s Law”, PLoS ONE **7**, 7, e40393, URL `http://dx.plos.org/10.1371/journal.pone.0040393` (2012).

Graedel, T. E., “Industrial ecology and the ecocity”, The Bridge **29**, 4, 10–14 (1999).

Hall, P., *Cities in Civilization* (Fromm Intl, Mount Prospect, IL, 1976).

Hall, P., *Cities in Civilization* (Pantheon Books, New York, 1998).

Harte, J., *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics* (Oxford University Press, New York, 2011).

Hausmann, R., “Housing versus Habitat”, URL `http://www.project-syndicate.org/commentary/reframing-low-income-housing-policy-by-ricardo-hausmann`, [Online; posted 26-SEP-2013] (2013).

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

Hidalgo, C. A. and R. Hausmann, “The building blocks of economic complexity”, PNAS **106**, 25, 10570–10575 (2009).

Hidalgo, C. A., B. Klinger, A.-L. Barabasi and R. Hausmann, “The product space conditions the development of nations”, Science **317**, 5837, 482–487 (2007).

Ibragimov, R. and S. Sharakhmetov, “The best constant in the rosenthal inequality for non-negative random variables”, Statistics & probability letters **55**, 4, 367–376 (2001).

Ibragimov, R. and S. Sharakhmetov, “The exact constant in the Rosenthal inequality for random variables with mean zero”, Theory of Probability & Its Applications **46**, 1, 127–132 (2002).

Ioannides, Y. and S. Skouras, “US city size distribution: Robustly Pareto, but only in the tail”, Journal of Urban Economics **73**, 1, 18–29, URL `http://www.sciencedirect.com/science/article/pii/S0094119012000447` (2013).

Jacobs, J., *The Economy of Cities* (Vintage Books, New York, 1969).

Jacobs, J., *Cities and the Wealth of Nations* (Vintage Books, New York, 1985).

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

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

Jones, C. I. and P. M. Romer, “The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital”, American Economic Journal: Macroeconomics **2**, 224–245 (2010).

Katz, J. S., “Indicators for complex innovation systems”, Research Policy **35**, 7, 893–909 (2006).

Kleiber, C. and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley Series in Probability and Statistics (John Wiley & Sons, 2003).

Klimek, P., R. Hausmann and S. Thurner, “Empirical confirmation of creative destruction from world trade data”, PLoS ONE **7**, 6, e38924, URL `http://dx.doi.org/10.1371%2Fjournal.pone.0038924` (2012).

Klimek, P., S. Thurner and R. Hanel, “Evolutionary Dynamics from a Variational Principle”, Physical Review E **82**, 1, 011901 (2010).

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

Lin, Z. and Z. Bai, *Probability Inequalities* (Springer, 2011), URL `http://books.google.com/books?id=Cxq1Qk7fiSAC`.

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

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

Malevergne, Y., V. Pisarenko and D. Sornette, "Testing the Pareto against the lognormal distribution with the uniformly most powerful unbiased test applied to the distribution of cities", Physical Review E **83**, 3, 036111(11) (2011).

Mandelbrot, B., "Post scriptum to "final note"", Information and Control **4**, 23, 300–304, URL `http://www.sciencedirect.com/science/article/pii/S0019995861800259` (1961).

Mantegna, R. N. and H. E. Stanley, "Stochastic Process with Ultraslow Convergence to a Gaussian: the Truncated Lévy Flight", Physical Review Letters **73**, 22, 2946 (1994).

Mantovani, M. C., H. V. Ribeiro, E. K. Lenzi, S. Picoli and R. S. Mendes, "Engagement in the electoral processes: Scaling laws and the role of political positions", Phys. Rev. E **88**, 024802, URL `http://link.aps.org/doi/10.1103/PhysRevE.88.024802` (2013).

Marin, J.-M. and C. Robert, *Bayesian Core: A Practical Approach to Computational Bayesian Statistics: A Practical Approach to Computational Bayesian Statistics* (Springer, 2007).

Marx, M., D. Strumsky and L. Fleming, "Mobility, Skills, and the Michigan Non-Compete Experiment", Management Science **55**, 6, 875–889 (2009).

Mayo, D. G., *Error and the Growth of Experimental Knowledge* (University of Chicago Press, 1996).

Mehta, N. B., J. Wu, A. F. Molisch and J. Zhang, "Approximating a sum of random variables with a lognormal", Wireless Communications, IEEE Transactions on **6**, 7, 2690–2699 (2007).

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

Mitzenmacher, M., "A Brief History of Generative Models for Power Law and Lognormal Distributions", Internet Mathematics **1**, 2, 226–251 (2003).

Montroll, E. W. and M. F. Shlesinger, "On $1/f$ noise and other distributions with long tails", Proc. Natl. Acad. Sci. USA **79**, 3380–3383, applied Mathematical Sciences (1982).

Moretti, E., *The New Geography of Jobs* (Houghton Mifflin Harcourt, 2012).

Mumford, L., *The City in History: Its Origins, Its Transformations, and Its Propects* (Harcourt Brace & World Inc, New York, NY, 1961).

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

Neffke, F., M. Hartog, R. Boschma and M. Henning, "Agents of structural change", Unpublished manuscript. `http://www.frankneffke.com/files/NeffkeHartog_etal2013.pdf` (2013).

Neffke, F. and M. Henning, “Skill relatedness and firm diversification”, Strategic Management Journal **34**, 3, 297–316 (2013).

Newman, M. E. J., “Power laws, Pareto distributions and Zipf’s law”, Cont. Phys. **46**, 5, 323–351 (2005).

Ó hUallacháin, B., “Patent places: Size matters”, Journal of Regional Science **39**, 4, 613–636 (1999).

Olver, F. W. J., D. W. Lozier, R. F. Boisvert and C. W. Clark, *NIST Handbook of Mathematical Functions* (Cambridge University Press, http://dlmf.nist.gov/, 2010).

Ortman, S. G., A. H. F. Cabaniss, J. O. Sturm and L. M. A. Bettencourt, “The Pre-History of Urban Scaling”, PLoS ONE **9**, 2, 1–10, URL `http://search.ebscohost.com.ezproxy1.lib.asu.edu/login.aspx?direct=true&db=aph&AN=94729679&site=ehost-live` (2014).

O’Sullivan, A., *Urban Economics* (McGraw-Hill/Irwin, New York, NY, 2006), 6th edn.

Pickett, S., M. Cadenasso, J. Grove, C. G. Boone, P. M. Groffman, E. Irwin, S. S. Kaushal, V. Marshall, B. P. McGrath, C. Nilon, R. Pouyat, K. Szlavecz, A. Troy and P. Warren, “Urban ecological systems: Scientific foundations and a decade of progress”, Journal of Environmental Management **92**, 331–362 (2011).

Piketty, T. and E. Saez, “Income Inequality in the United States, 1913–1998”, The Quarterly Journal of Economics **118**, 1–39, updated information in `http://elsa.berkeley.edu/~saez/.` (2003).

Pinker, S., *The better angels of our nature: why violence has declined*, Viking (Penguin Group, New York, 2011).

Polèse, M., “Cities and National Economic Growth: A Reappraisal”, Urban Studies **42**, 8, 1429–1451 (2005).

Polèse, M., “Five Principles of Urban Economics”, City Journal **23**, 1, URL `www.city-journal.org/2013/23_1_urban-economics.html`, www.city-journal.org/2013/23_1_urban-economics.html (online article) (2013).

Portugali, J., *Self-Organization and the City* (Springer, 2000).

Portugali, J., H. Meyer, E. Stolk and E. Tan, *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design* (Springer, 2012).

Puga, D., “The Magnitude and Causes of Agglomeration Economies”, Journal of Regional Science **50**, 1, 203–219 (2010).

Quigley, J. M., “Urban Diversity and Economic Growth”, Journal of Economic Perspectives **12**, 2, 127–138 (1998).

Rauch, J. E., “Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities”, Journal of Urban Economics **34**, 3, 380–400, URL `http://www.sciencedirect.com/science/article/pii/S0094119083710429` (1993).

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

Romeo, M., V. Da Costa and F. Bardou, “Broad distribution effects in sums of lognormal random variables”, The European Physical Journal B **32**, 4, 513–525 (2003).

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

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

Rosenthal, H. P., “On the subspaces of $l^p (p > 2)$ spanned by sequences of independent random variables”, Israel Journal of Mathematics **8**, 3, 273–303 (1970).

Rosenthal, S. S. and W. C. Strange, “Evidence on the nature and sources of agglomeration economies”, Handbook of regional and urban economics **4**, 2119–2171 (2004).

Rosenthal, S. S. and W. C. Strange, “The micro-empirics of agglomeration economies”, A Companion to Urban Economics pp. 7–23 (2006).

Roy, A. D., “The Distribution of Earnings and of Individual Output”, The Economic Journal **60**, 239, 489–505 (1950).

Rozenfeld, H. D., D. Rybski, J. S. Andrade, M. Batty, H. E. Stanley and H. A. Makse, “Laws of population growth”, Proceedings of the National Academy of Sciences **105**, 48, 18702–18707 (2008).

Rozenfeld, H. D., D. Rybski, X. Gabaix and H. A. Makse, “The area and population of cities: New insights from a different perspective on cities”, Tech. rep., National Bureau of Economic Research (2009).

Sah, R. K., “Social Osmosis and Patterns of Crime”, Journal of Political Economy **99**, 6, 1272–1295, URL `http://www.jstor.org/stable/2937730` (1991).

Saichev, A., Y. Malevergne and D. Sornette, *Theory of Zipf’s Law and Beyond*, Lecture Notes in Economics and Mathematical Systems 632 (Springer-Verlag, Berlin Heidelberg, 2010).

Saw, J. G., M. C. K. Yang and T. C. Mo, “Chebyshev Inequality With Estimated Mean and Variance”, The American Statistician **38**, 2, 130–132 (1984).

Schroeder, M., *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*, Dover Books on Physics (Dover Publications, New York, 1991), 2009 dover edn.

Scott, A. J. and M. Storper, “The Nature of Cities: The Scope and Limits of Urban Theory”, International Journal of Urban and Regional Research p. (electronic version), URL `http://dx.doi.org/10.1111/1468-2427.12134` (2014).

Shalizi, C. R., “Scaling and hierarchy in urban economies”, arXiv preprint arXiv:1102.4101 (2011).

Shiryaev, A. N., *Probability*, Graduate Texts in Mathematics (Book 95) (Springer, New York, 1995), 2nd edn.

Shockley, W., "On the Statistics of Individual Variations of Productivity in Research Laboratories", Proceedings of the IRE **45**, 279–290 (1957).

Simon, H. A., "On a Class of Skew Distribution Functions", Biometrika **42**, 3/4, 425–440 (1955).

Simon, H. A., "On Judging the Plausibility of Theories", in "Logic, Methodology and Philosophy of Science III", edited by B. V. Rootselaar and J. Staal, vol. 52 of *Studies in Logic and the Foundations of Mathematics*, pp. 439–459 (Elsevier, 1968), URL `http://www.sciencedirect.com/science/article/pii/S0049237X08712114`.

Simon, M. A. and G. W. Blume, "Building and understanding multiplicative relationships: A study of prospective elementary teachers", Journal for Research in Mathematics Education pp. 472–494 (1994).

Soo, K. T., "Zipf's Law for cities: a cross-country investigation", Regional Science and Urban Economics **35**, 239–263, URL `http://www.sciencedirect.com/science/article/pii/S016604620400033X` (2005).

Sornette, D., *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*, Springer Series in Synergetics (Springer, Heidelberg, 2006), 2nd edn.

Sornette, D., "Probability Distributions in Complex Systems", in "Computational Complexity", pp. 2286–2300 (Springer, 2012).

Sornette, D. and R. Cont, "Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws", J. Phys. I France **7**, 431–444 (1997).

Sornette, D., T. Maillart and G. Ghezzi, "How much is the whole really more than the sum of its parts? $1 + 1 = 2.5$: Superlinear productivity in collective group actions", arXiv preprint arXiv:1405.4298 (2014).

Storper, M., C. van Marrewijk and F. G. van Oort, "Introduction: Processes of Change in Urban Systems", Journal of Regional Science **52**, 1, 1–9 (2012).

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

Sveikauskas, L., "The Productivity of Cities", The Quarterly Journal of Economics **89**, 3, 393–413, URL `http://www.jstor.org/stable/1885259` (1975).

Szyszkowicz, S. S. and H. Yanikomeroglu, "Modified-power-lognormal approximation to the sum of lognormals distribution", in "Global Telecommunications Conference", pp. 1–6 (Citeseer, 2009).

The World Bank, “Conflict, Security, and Development”, Report, The International Bank for Reconstruction and Develpment/The World Bank, 1818 H Street NW, Washington DC 20433 (2011a).

The World Bank, “Violence in the City: Understanding and Supporting Community Responses to Urban Violence”, Report, The International Bank for Reconstruction and Develpment/The World Bank, 1818 H Street NW, Washington DC 20433 (2011b).

The World Bank, “Global Monitoring Report 2013: Rural-Urban Dynamics and the Millennium Development Goals”, Report, The International Bank for Reconstruction and Develpment/The World Bank, 1818 H Street NW, Washington DC 20433, the World Bank and the International Monetary Fund (2013).

Thurner, S., P. Klimek and R. Hanel, “Schumpeterian Economic Dynamics as a Quantifiable Model of Evolution”, New Journal of Physics **12**, 7, 075029 (2010).

United Nations, “World Economic and Social Survey 2013: Sustainable Development Challenges”, Report, Department of Economic and Social Affairs, New York, e/2013/50/Rev. 1 (2013).

Uslaner, E. M., “The Pitfalls of Per Capita”, American Journal of Political Science **20**, 1, 125–133 (1976).

Valle-Jones, D., *Violent Deaths in Mexico 2004-2009*, URL `https://github.com/diegovalle/death.index`, r package version 0.0.1 (2011).

Warren, P., S. Harland, C. Boone, S. B. Lerman, E. Shochat and A. P. Kinzig, *Urban Ecology* (Cambridge University Press, London, 2010).

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

West, G. B. and J. H. Brown, “Review: The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization”, The Journal of Experimental Biology **208**, 1575–1592 (2005).

West, G. B., J. H. Brown and B. J. Enquist, “A General Model for the Origin of Allometric Scaling Laws in Biology”, Science **276**, 122–126 (1997).

West, G. B., J. H. Brown and B. J. Enquist, “The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms”, Science **284**, 1677–1679 (1999).

Whitfield, J., *In the Beat of a Heart: Life, Energy, and the Unity of Nature* (Joseph Henry Press, Washington, D.C., 2006).

Youn, H., L. M. A. Bettencourt, J. Lobo, D. Strumsky, H. Samaniego and G. B. West, “The systematic structure and predictability of urban business diversity”, Papers, arXiv.org, URL `http://EconPapers.repec.org/RePEc:arx:papers:1405.3202` (2014).

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

APPENDIX A

DATA

A.1 Socio-Economic Productivity

The source of the following data is the U.S. Bureau of Economic Analysis (BEA).

A.1.1 Real Personal Income

1. Go to the interactive tables of the BEA regarding the Regional Data on GDP & Personal Income.
2. In the box "REAL PERSONAL INCOME AND REGIONAL PRICE PARITIES", click on "Real personal income (RPI1)".
3. Mark "Metropolitan Statistical Area" and press "Next Step".
4. Highlight "All Areas" in the Area box, select "Levels" as the Unit Of Measure and press "Next Step".
5. Select the years of interest and press "Next Step".

A.1.2 Real GDP Per Capita

1. Go to the interactive tables of the BEA regarding the Regional Data on GDP & Personal Income.
2. In the box "GROSS DOMESTIC PRODUCT BY METRO AREA", click on "Real per capita GDP".
3. Select "All MSAs", select "Levels" as the Unit Of Measure and press "Next Step".
4. Select the years of interest and press "Next Step".

A.1.3 Real Wages

$$\begin{aligned} \text{Real wages per job} &= \frac{\text{Real wages}}{\text{Number of jobs}} \\ &= \frac{\text{Wages and salaries}/\text{Implicit regional price deflator index}}{\text{Number of jobs}}. \end{aligned} \tag{A.1}$$

A.2 Homicides

Homicides are defined as deaths caused by other persons, intentionally or not. Data for Colombia is available online at the National Institute of Legal Medicine and Forensic Sciences (`http://www.medicinalegal.gov.co`) and municipality populations at the National Administrative Department of Statistics (`http://www.dane.gov.co`). Brazil's population and homicide numbers are available from the Sangari Institute and Brazilian Ministry of Justice (`http://www.sangari.com/mapadaviolencia/`). The data spans the years 2003-2007 for Brazil, 2004-2009 for Colombia, and 2005-2009 for Mexico. Data for Mexican municipalities was compiled by Valle-Jones (2011).

We adopted standard definitions of metropolitan areas available at http://www.secretariasenado.gov.co/senado/basedoc/ley/1994/ley_0128_1994.html for Colombia. However, comprehensive definitions for many metropolitan areas in Colombia do not exist officially although they are recognized in various contexts (see for example http://www.dane.gov.co/files/censo2005/resultados_am_municipios.pdf for the case of Bogotá). We adopted such unofficial definitions in our analysis. For Mexico definitions are available at the National Institute of Statistics and Geography, http://www.inegi.gob.mx/est/contenidos/espanol/metodologias/otras/zonas_met.pdf, and for Brazil at the Observatory of the Metropolis, http://www.observatoriodasmetropoles.ufrj.br/metrodata/ibrm/index.html.

APPENDIX B

A QUESTION ABOUT MARKOV-TYPE INEQUALITIES

Markov-type inequalities impose bounds on tail probabilities over a random variable X when we have knowledge about the moments of its distribution. Markov's inequality, specifically states the following:

Theorem (Markov's inequality)**.** *Let X be a non-negative random variable and $t > 0$. Then,*

$$\Pr\{X \geq t\} \leq \frac{\mathrm{E}[X]}{t}. \tag{B.1}$$

This theorem and others like it (e.g., Chebyshev's inequality) are of little practical value. If one knows the distribution of X then one can compute both $\Pr\{X \geq t\}$ and $\mathrm{E}[X]$. The value of the theorem is when one knows $\mathrm{E}[X]$ but does not know the distribution. Note, however, that in practice we seldom have knowledge about $\mathrm{E}[X]$. Instead, one usually has an estimation $\bar{x}$ of $\mathrm{E}[X]$.

Assuming a sequence of random variables $X_1, X_2, \ldots, X_n$ identically distributed to X, and their average $X^{(n)} \equiv \bar{x} = \sum_{i=1}^{n} X_i/n$, one would like a statement that provides bounds on $\Pr\{X \geq t\}$.

In the inequality (B.1) both sides are numbers. Asking for an analogous inequality when one does not have knowledge about $\mathrm{E}[X]$ but rather about $X^{(n)}$ requires the right-hand side of equation (B.1) to become a random variable (see, however, Saw *et al.*, 1984). Hence, such inequality has to be framed as a probabilistic statement. Thus, one has the following question: How is

$$\Pr\left\{\Pr\{X \geq t\} \leq \frac{X^{(n)}}{t}\right\}$$

bounded?

Rearranging,

$$\Pr\left\{\; X^{(n)} \geq t\ \Pr\{X \geq t\} \;\right\} \;\leq\; ? \tag{B.2}$$

One cannot use Markov's inequality again (replacing $X \to X^{(n)}$ and $t \to t\Pr\{X \geq t\}$). Since $\mathrm{E}\left[X^{(n)}\right] = \mathrm{E}[X]$, the bound becomes trivially 1.[1]

This question is related to the one addressed in chapter 2. This is because the sequence $X_1, \ldots, X_n$ can be regarded as the productivity of all n individuals living in a certain city. Since what we often know is the aggregate productivity $S_n = \sum_{i=1}^{n} X_i$, the question is what are the constraints on the probability distribution of X_i from knowing S_n?

[1]From Markov's inequality, $1 \leq \frac{\mathrm{E}[X]}{t\Pr\{X > t\}}$.

APPENDIX C

SCALING OF THE 'LEVY CASE'

We follow an intuitive proof of equation (2.3) borrowed from Sornette (2006)[1]:

$$\begin{aligned} \mathrm{E}[S_N] &= N\,\mathrm{E}[X_1] \\ &= N\int_{x_0}^{\infty} x p_X(x)\mathrm{d}x. \end{aligned} \tag{C.1}$$

As was mentioned in section 2.2.1, the integral in equation (C.1) diverges. For a fixed sample size N, however, there will always be a maximum $M_N = \max\{X_1, \ldots, X_N\}$. This maximum allows one to truncate the integral.

The random variable M_N corresponding to the maximum in the sample has itself an associated probability density function $p_{M_N}(x)$, which we must approximate in order to be able to compute the integral in equation (C.1). One can find an approximation of this function noting that the probability that $\Pr\{M_N < x\} = \Pr\{X_1 < x\} \cdot \Pr\{X_2 < x\} \cdots \Pr\{X_N < x\} = (\Pr\{X_1 < x\})^N$. In other words, the probability that M_N is less than x is equal to the product (remember we assume independence) of the probabilities that each of the N random variables X_i is less than x. Rearranging,

$$\begin{aligned} \Pr\{M_N < x\} &= (\Pr\{X_1 < x\})^N, \\ &= (1 - \Pr\{X_1 \geq x\})^N, \\ &= \exp\{N \ln[1 - \Pr\{X_1 \geq x\}]\}. \end{aligned} \tag{C.2}$$

For large values of x the probability $\Pr\{X_1 \geq x\}$ is very small, so we can expand the logarithm using $\ln(1-x) \approx -x$:

$$\Pr\{M_N < x\} \approx \exp\{-N \Pr\{X_1 \geq x\}\}. \tag{C.3}$$

Given this, we can compute the value $x^{(p,N)}$ which the maximum will not exceed with probability $p \equiv \Pr\left\{M_N < x^{(p,N)}\right\}$:

$$p \approx \exp\{-N \Pr\left\{X_1 \geq x^{(p,N)}\right\}\}, \tag{C.4}$$

so

$$\Pr\left\{X_1 \geq x^{(p,N)}\right\} \approx \frac{\ln(1/p)}{N}. \tag{C.5}$$

For the case of a power-law probability distribution such as equation (2.2), the quantile value $x^{(p,N)}$ of the maximum is calculated from

$$\left(\frac{x^{(p,N)}}{x_0}\right)^{-\tau} \approx \frac{\ln(1/p)}{N}, \tag{C.6}$$

whose solution yields

$$x^{(p,N)} \approx C_p\, N^{\frac{1}{\tau}}, \tag{C.7}$$

[1] We refer the reader to Bouchaud and Georges (1990) for a more thorough and detailed discussion of this and related topics.

where

$$C_p = \left[\frac{x_0^\tau}{\ln(1/p)}\right]^{\frac{1}{\tau}}. \tag{C.8}$$

Using equation (C.7) in equation (C.1) to truncate the integral, we can say that, with probability p,

$$\begin{aligned}
\mathrm{E}\left[S_N\right] &\approx N \int_{x_0}^{x^{(p,N)}} x p_X(x) \mathrm{d}x \\
&= N \int_{x_0}^{x^{(p,N)}} x \left(\frac{\tau}{x_0}\right) \left(\frac{x}{x_0}\right)^{-\tau-1} \mathrm{d}x \\
&= N\tau \int_{x_0}^{x^{(p,N)}} \left(\frac{x}{x_0}\right)^{-\tau} \mathrm{d}x \\
&= N\tau x_0^\tau \frac{\left[\left(x^{(p,N)}\right)^{-\tau+1} - x_0^{-\tau+1}\right]}{-\tau+1} \\
&\approx \frac{N\tau x_0^\tau}{1-\tau} \left[\left(C_p\, N^{\frac{1}{\tau}}\right)^{-\tau+1} - x_0^{-\tau+1}\right] \\
&= \frac{\tau\, x_0}{1-\tau} \left[(\ln(1/p))^{1-1/\tau} N^{1/\tau} - N\right] \\
&\propto N^{1/\tau},
\end{aligned} \tag{C.9}$$

for $\tau < 1$ and large N.

APPENDIX D

KESTEN PROCESS

In chapter 2 we assumed that the distribution of productivities X_i was a lognormal with heavy tails (i.e., $\sigma^2 \geq 1$). This distribution allowed us to show that even when all the moments of a distribution are finite, the sum of $\sum_{i=1}^{N} X_i$ scales superlinearly with population size N (fig. 2.3), and the convergence to the Law of Large Numbers is very slow (fig. 2.6). A concern, however, is whether the assumption of lognormality is reasonable. In this section we discuss briefly a simple model that justifies this assumption (Gabaix, 2009).

For the sake of concreteness, let $X_{i,t}$ represent the total wage (i.e., productivity) of individual i at time t. The model is constructed upon two ideas:

1. Individuals receive in each time period a random quantity $w_{i,t} > 0$ (think of it as a continual influx of money);

2. Their current wage is their previous wage affected by random multiplicative shocks, denoted by $m_{i,t}$.

Given this, the total wage of individual i in the period $t+1$ is given by the following recursive equation:

$$X_{i,t+1} = m_{i,t} X_{i,t} + w_{i,t}, \tag{D.1}$$

where the subscript i will be dropped for simplicity in what follows.

The multiplicative term m_t in equation (D.1) is meant to represent the advantages ($m_t > 1$), or disadvantages ($m_t < 1$), of urban interactions. Social systems, including spatial agglomerations and cities, are axiomatically characterized by an intricate web of interactions. As a consequence, the individuals in the system acquire and develop their properties not only from their inherent endowments but more importantly, from their interaction with their social and physical environments. Mathematically, the interaction between two variables is typically represented in a model in the form of a multiplicative term (Aiken and West, 1991).[1]

Equation (D.1) generates a probability density function $p_X(x) \propto x^{-1-\nu}$ for a wide variety of distribution functions for m_t and a_t, such that $m_t > 0$, $a_t > 0$, and $\langle \ln(a) \rangle < 0$. More generally, simple multiplicative processes such as these will generate variables with distributions like lognormals (when $w_t = 0$ and $X_0 > 0$), stretched exponentials, and power-laws (Sornette and Cont, 1997).

Figure D.1 plots the distribution of a particular simulation of the model. There, the wage of $10,000$ individuals has been simulated according to equation (D.1), where each time step represents a year, and the wage of each individual i was simulated T_i time steps (representing the lifespan of an individual). Here, the T_i were generated from a normal distribution with mean 30 and standard deviation 20, and truncated so that $T_i > 0$ for all i.

This simple model exemplifies the broad distributions that result from multiplicative processes. We refer the reader to the results by Sornette and Cont (1997), which provide a stochastic analytical framework to understand the origin of the broad distributions that might describe the socio-economic properties of individuals in a city.

[1]Intrinsic difficulties in understanding the meaning of multiplicative relationships has been reported in the psychological literature, with interesting consequences for pedagogy (see, e.g., Simon and Blume, 1994).

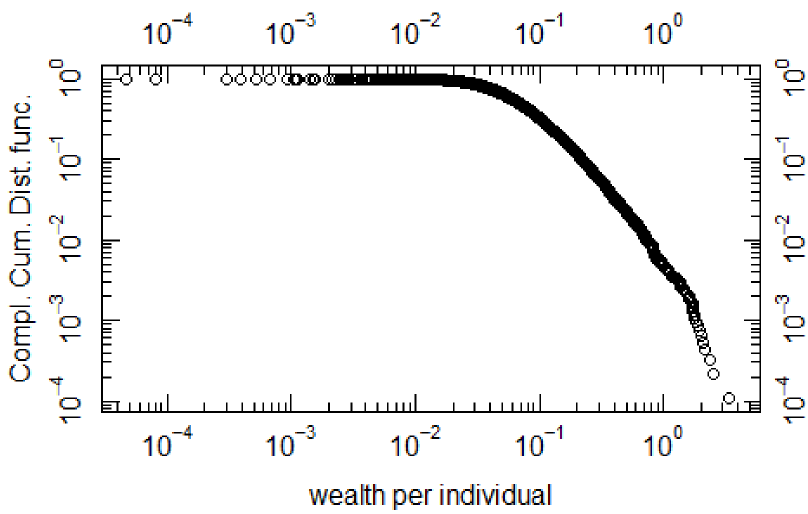


Figure D.1: Distribution of Wage in Kesten Process. Simulated from 10,000 Individuals Using Equation (D.1) with Distributions of Age from a Truncated Normal Distribution.

APPENDIX E

COMBINING MARKOV’S AND ROSENTHAL’S INEQUALITIES

The following inequalities we use to find the interplay between the population size N and the moments of the distribution of $X_1,\ldots,X_N$ that constitute the sum $Y_N=\sum_{i=1}^N X_i$:

Theorem (Markov's generalized inequality)**.** *Let h be a non-negative, non-decreasing function and let X be a non-negative random variable. Then*

$$\Pr\{X\geq t\}\leq\frac{\mathrm{E}\left[h(X)\right]}{h(t)}, \tag{E.1}$$

for any $t>0$.

The other inequality we will use is the Rosenthal inequality (see Rosenthal, 1970 or Lin and Bai, 2011, Thms. 9.7.b & 9.7.c):

Theorem (Rosenthal's inequality 2)**.** *Let $X_1,X_2,\ldots,X_N$ be a sequence of i.i.d. random variables, with $\mathrm{E}\left[X_1\right]=0$, $\mathrm{E}\left[|X_1|^r\right]<\infty$, and let $Y_N=\sum_{i=1}^N X_i$. Then*

$$A(r)\max\left\{N\mathrm{E}\left[|X_1|^r\right],\left(N\mathrm{E}\left[X_1^2\right]\right)^{r/2}\right\}\leq\mathrm{E}\left[|Y_N|^r\right]\leq c(r)\max\left\{N\mathrm{E}\left[|X_1|^r\right],\left(N\mathrm{E}\left[X_1^2\right]\right)^{r/2}\right\}, \tag{E.2}$$

for some constants $A(r)$ and $c(r)$ only dependent on $2\leq r<\infty$.

E.1 Derivation

Recall that $X_1,\ldots,X_N$ are i.i.d. r.v.'s with $\mathrm{E}\left[X_1\right]=\mu$ and $Y_N=\sum_{i=1}^N X_i$.
Using Markov's inequality for $r>0$,

$$\Pr\left\{\left|\frac{Y_N}{N}-\mu\right|\geq\varepsilon\right\}\leq\frac{\mathrm{E}\left|\frac{Y_N}{N}-\mu\right|^r}{\varepsilon^r}. \tag{E.3}$$

Let $Z_i\equiv\frac{X_i-\mu}{N}$ and let $S_N=\sum_{i=1}^N Z_i=\frac{Y_N}{N}-\mu$. Since Z_i are i.i.d. with $\mathrm{E}\left[Z_i\right]=0$, Rosenthal's inequality states that if $\mathrm{E}|Z_1|^r<\infty$, then

$$\mathrm{E}|S_N|^r\leq c(r)\max\left\{N\mathrm{E}|Z_1|^r,\left(N\mathrm{E}\left[Z_1^2\right]\right)^{r/2}\right\}, \tag{E.4}$$

for a constant $c(r)$ and $2\leq r<\infty$.
Combining the last two equations, we get

$$\Pr\left\{\left|\frac{Y_N}{N}-\mu\right|\geq\varepsilon\right\}\leq\frac{c(r)\max\left\{N\mathrm{E}|Z_1|^r,\left(N\mathrm{E}\left[Z_1^2\right]\right)^{r/2}\right\}}{\varepsilon^r}. \tag{E.5}$$

In order for this inequality to make sense, the right-hand side must be less than 1, and since we want the left-hand side to be at worst $\alpha>0$, this becomes

$$\alpha\geq\frac{c(r)\max\left\{N\mathrm{E}|Z_1|^r,\left(N\mathrm{E}\left[Z_1^2\right]\right)^{r/2}\right\}}{\varepsilon^r}. \tag{E.6}$$

Noting that $\mathrm{E}|Z_1|^r = \frac{1}{N^r}\mathrm{E}|X_i - \mu|^r$ and $\mathrm{E}\left[Z_1^2\right] = \frac{1}{N^2}\mathrm{Var}[X_1]$, we get

$$(\alpha\varepsilon^r)N^r \geq c(r)\left(N\mathrm{E}|X_1 - \mu|^r + N^{r/2}(\mathrm{Var}[X_1])^{r/2}\right). \tag{E.7}$$

APPENDIX F

ACCOMPANYING PLOTS FOR HISTOGRAMS OF THE POISSON MODEL

F.1 Parameter Estimates

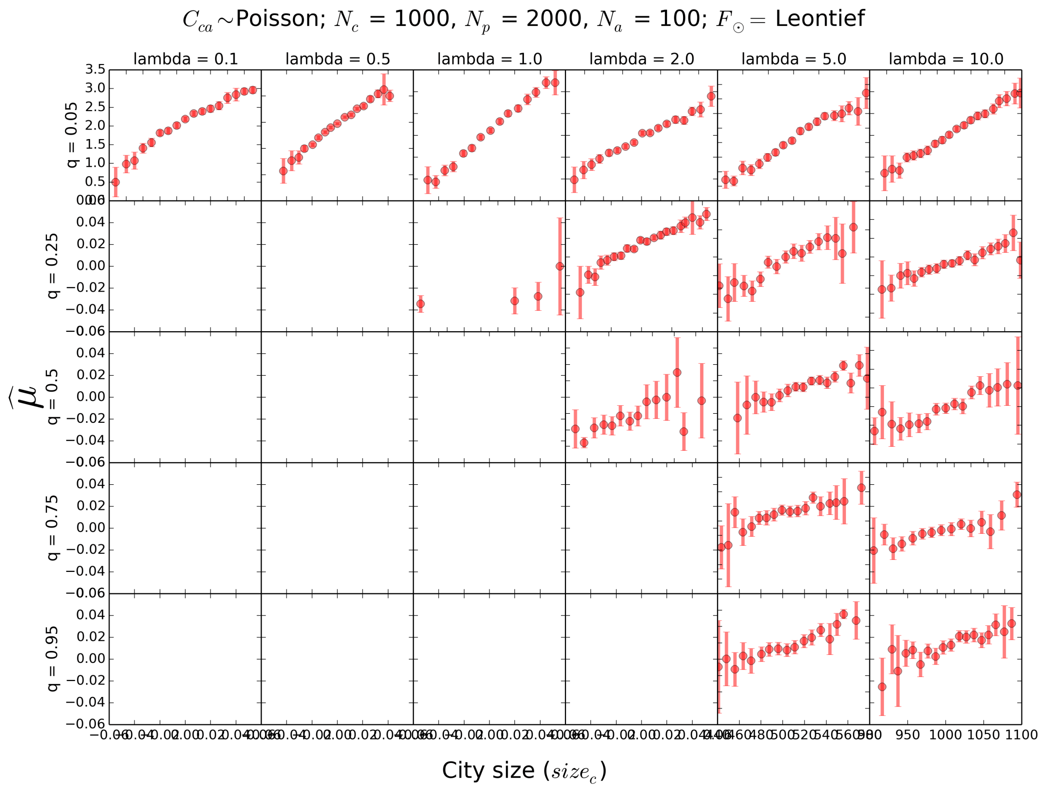


Figure F.1: Estimates of the μ Parameter of the Distribution of Standardized $\ln y_c$ Conditional on $size_c$, Binned Linearly, for Different Combinations of the Parameters q and λ. Here, $N_c = 1000, N_p = 2000, N_a = 100$.

F.2 Other Production Functions

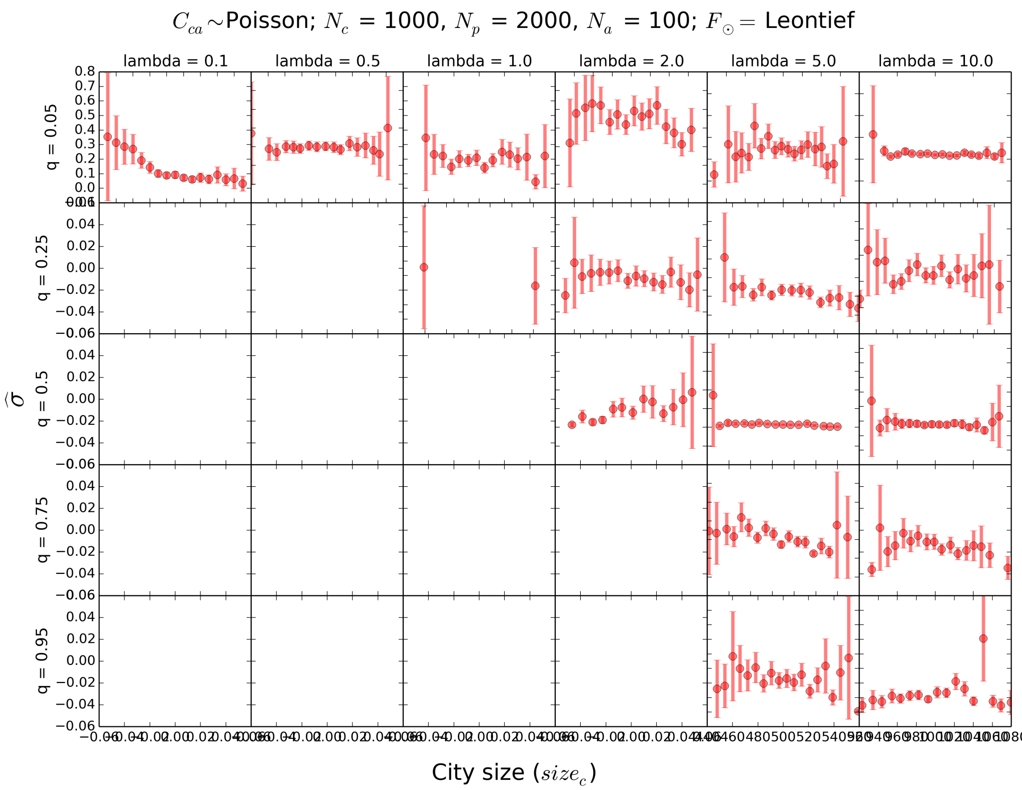


Figure F.2: Estimates of the σ Parameter of the Distribution of Standardized $\ln y_c$ Conditional on $size_c$, Binned Linearly, for Different Combinations of the Parameters q and λ. Here, $N_c = 1000, N_p = 2000, N_a = 100$.

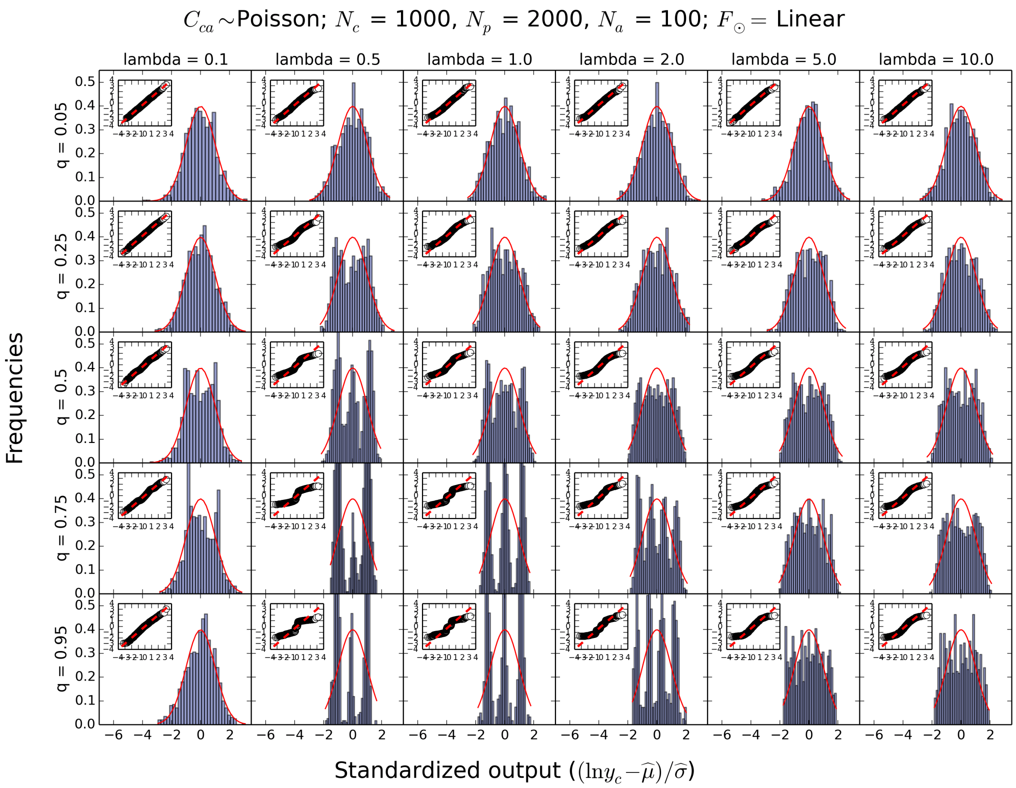


Figure F.3: Histograms, Given a Linear Production Function, of Standardized $\ln y_c$ Conditional on $size_c$ (Binned Linearly) for Different Combinations of the Parameters q and λ. The Inset Plots Are Q-Q Plots for the Normal Distribution. If the Output is Lognormally Distributed the Dots Should Line Up with the Red Dashed Line. Here, $N_c = 1000, N_p = 2000, N_a = 100$.

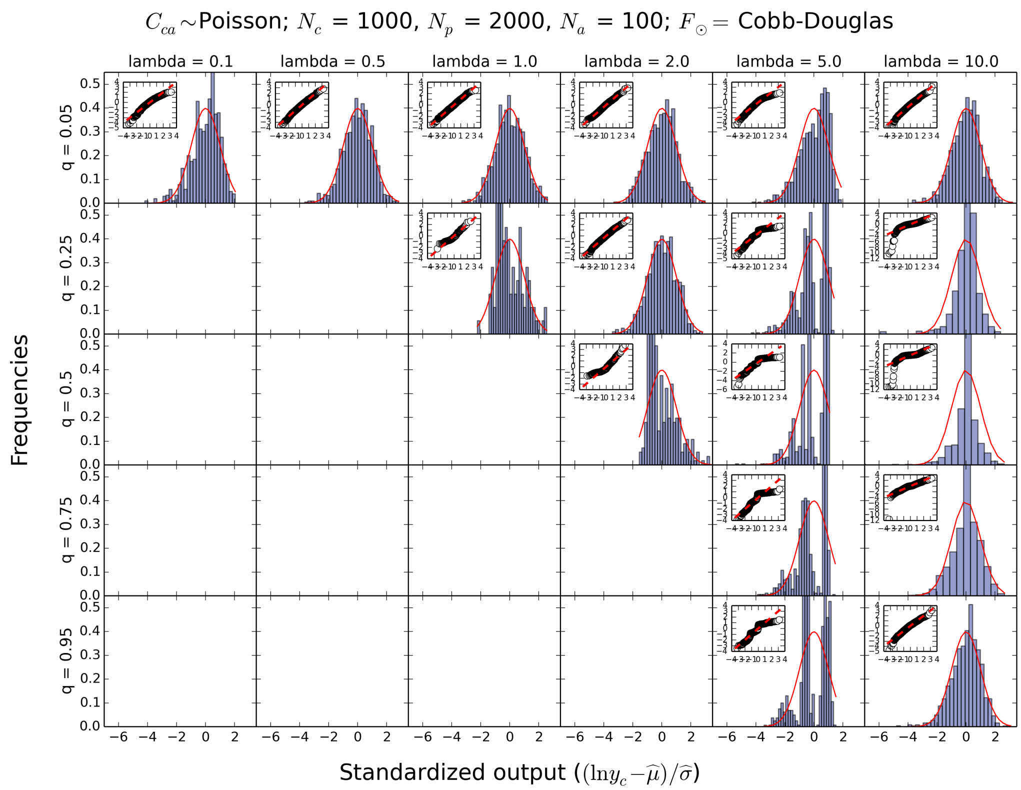


Figure F.4: Histograms, Given a Cobb-Douglas Production Function, of Standardized $\ln y_c$ Conditional on $size_c$ (Binned Linearly) for Different Combinations of the Parameters q and λ. The Inset Plots Are Q-Q Plots for the Normal Distribution. If the Output is Lognormally Distributed the Dots Should Line Up with the Red Dashed Line. Here, $N_c = 1000, N_p = 2000, N_a = 100$.

APPENDIX G

SUPPORTING INFORMATION FOR CREATIVE AND INVENTIVE ACTIVITIES

Table G.1: 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**

Table G.2: 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. 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**

G.1 Conditional Distributions Goodness-Of-Fit Tests

In table G.1 and table G.2 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. 4.3 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.

G.2 Marginal Distribution Fits

The rationale behind equation (4.4) 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 (fig. 4.1) 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.

Figure G.1 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

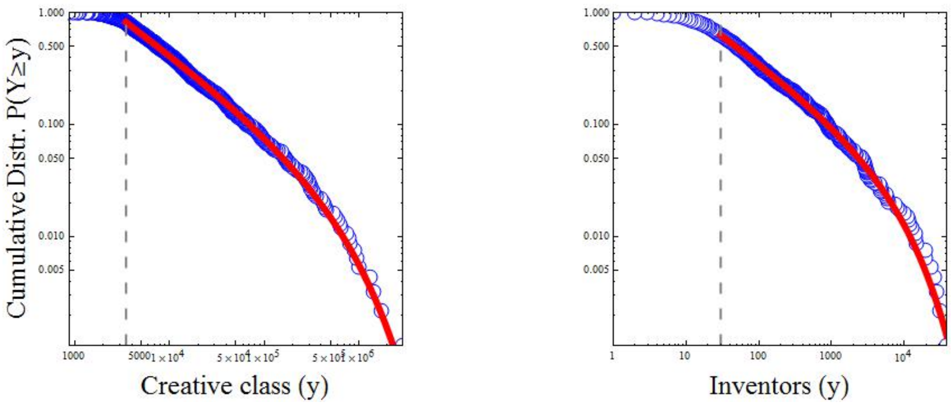


Figure G.1: Marginal Distributions of Creatives C and Inventors I Are Well Fit by a Power-Law With an Exponential Cutoff. We Plot the Empirical Complementary Cumulative Distribution (Blue Circles) with the Fit (Red Solid Line) Corresponding to the Function in equation (G.1) Using Maximum Likelihood. 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.

characterization of this distribution is a power-law with an exponential cutoff:

$$p_Y(y;\tau,\gamma,y_{\min}) = \frac{\gamma^{\tau-1}}{\Gamma(1-\tau, y_{\min}/\gamma)} \frac{\mathrm{e}^{-y/\gamma}}{y^{\tau}}, \quad y \geq y_{\min}, \tag{G.1}$$

where $\tau > 0$ 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[1]

$$\Gamma(z,a) = \int_a^{\infty} t^{z-1}\mathrm{e}^{-t}\mathrm{d}t. \tag{G.2}$$

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.

We implement the methodology presented in Clauset *et al.* (2009) to fit equation (G.1) to the data. We maximize the log-likelihood of the data for a given $y_{\min}$ above which the distribution holds. This minimum value is estimated as the one that minimizes the Kolmogorov-Smirnov distance between the empirical and the fitted distributions.

Now, from fig. 4.1B 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

$$p_Y(y)\mathrm{d}y = p_N(n)\mathrm{d}n \tag{G.3}$$

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

Table G.3: Different Estimated Probability Density Functions to Fit Population Size Distribution. PLEC Stands for "Power-Law with Exponential Cutoff", Given by equation (G.1), and PL Stands for "Power-Law", Given by a Density $p(n) \propto n^{-\alpha}$. All Four Distribution were Truncated from Below by the Same $\widehat{n_{\min}}$ Estimated Using equation (G.8).

Goodness-of-fit	$p_N(\alpha,\beta,\nu,n_{\min})$	PLEC($\tau,\gamma,n_{\min}$)	Lognormal(μ, σ)	PL($n_{\min}$, α)
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**

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

$$\begin{aligned} p_N(n) &= p_Y(y)\frac{\mathrm{d}y}{\mathrm{d}n} \\ &= \frac{\beta\gamma^{\tau-1}}{\Gamma(1-\tau, y_{\min}/\gamma)} \frac{\mathrm{e}^{-Y_0 n^\beta/\gamma}}{Y_0^{\tau-1} n^{\beta\tau-\beta+1}}. \end{aligned} \tag{G.4}$$

Equation (G.4) can be written as

$$p_N(n;\alpha,\beta,\nu,n_{\min}) = C\frac{\mathrm{e}^{-\left(\frac{n}{\nu}\right)^\beta}}{\left(\frac{n}{\nu}\right)^\alpha}, \quad n \geq n_{\min}, \tag{G.5}$$

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

$$\alpha = \beta(\tau-1)+1 \tag{G.6}$$

$$\nu = \left(\frac{\gamma}{Y_0}\right)^{1/\beta} \tag{G.7}$$

$$n_{\min} = \left(\frac{y_{\min}}{Y_0}\right)^{1/\beta}. \tag{G.8}$$

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 $\widehat{y_{\min}}$, from creative employment, we get from equations (G.6)-(G.8) that $\widehat{\alpha} = 1.675$, $\widehat{\beta} = 1.083$, $\widehat{\nu} = 11,329,658$, and $\widehat{n_{\min}} = 35,141.6$, which are the estimations reported in the main text.

Table G.3 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.

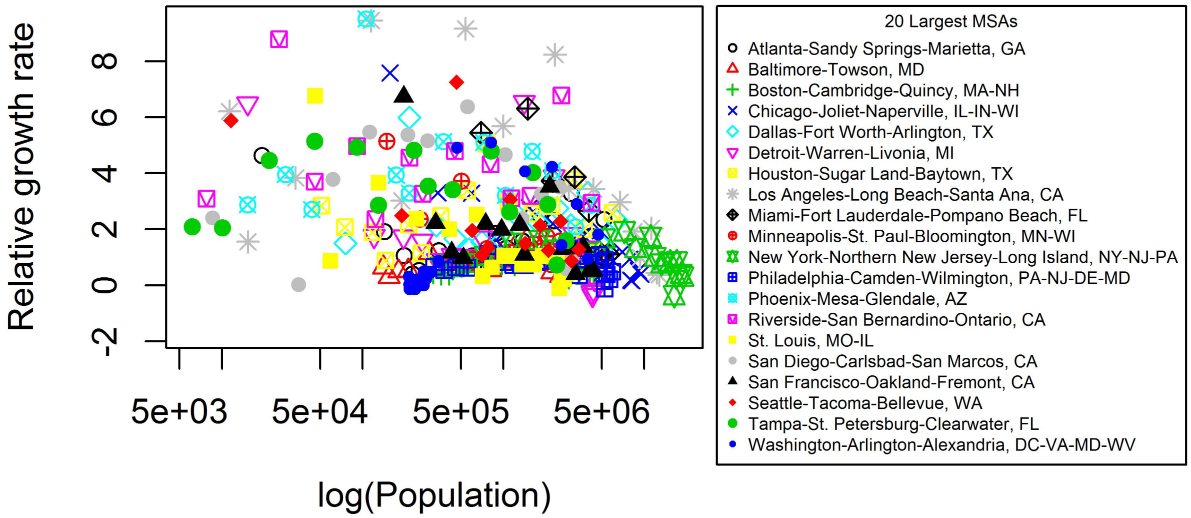


Figure G.2: Relative Growth Rates of the Twenty Largest MSAs as of 2010 Versus Their Population Size. The Data Goes Back to 1800 and the Data is from the Available Decennial U.S. Population Census. The Plot Suggest that when Cities Approach Population Sizes of Approximately 5 Million Inhabitants, their Growth Falls Relative to the Rest of the Country.

G.3 Population Growth Rate Versus Population Size

We plot in fig. G.2 the relative population growth rates per decade (i.e. the growth rate of city i divided by the national growth rate over ten year periods) versus the population size, for the 2010's twenty largest MSAs. As is custom the x-axis is plotted in logarithmic scale given the broad range of city populations observed. Although very few MSAs grow to the point of approaching our estimated value $\widehat{v} \approx 11,000,000$, a qualitative tendency is seen whereby the growth decays as cities become bigger.

APPENDIX H

REJECTING POISSON FOR THE CONDITIONAL DISTRIBUTION OF HOMICIDES

A maximum-likelihood procedure is presented here aimed to test whether the annual number of homicides in cities of a given size is described by a Poisson distribution. Homicides are assumed to be improbable events for which, when aggregated in annual counts, the law of rare events should hold. When we test the likelihood of this null hypothesis, we find that we can reject it with high confidence. This means that the underlying dynamics of homicides are more complex than the simple model implied by a Poisson process. The methodology is general enough to be applied to other distributions. Future work would also point to the development of a process by which crime is contagious through a topology of influences, which would introduce positive effects on the distribution of homicides.

In chapter 3 we found that: when Y represents the annual number of homicides and N the population size (both taken as random variables), we have empirically found that the probabilistic expression $P(N=n|Y=y)$ resembles a lognormal distribution, with $\mathbb{E}[N|Y=y]=N_0\ y^{1/\beta}$ and variance $Var(\ln N|Y=y)=\sigma^2$ independent of y (see fig. 3.4). When $P(Y=y)$ is power-law distributed, it can be shown that $P(Y=y|N=n)$ is also a lognormal distribution (using Bayes' rule).

The result about the lognormality of the conditional distribution of homicides given population size is surprising because the law of rare events would be expected to be at work. When dealing with accumulated annual number of homicides in a city, the distribution should result in a Poisson distribution instead of a lognormal distribution. In other words, the Poisson distribution is a null hypothesis that must be rejected before proposing other distributions.

Here we develop a procedure to test the hypothesis of a Poisson distributed number of homicides given a certain city population. We approach the problem using a maximum-likelihood method.

H.1 Methods

Let N and Y be the random variables representing the city population and the corresponding number of homicides in a year, and let n_i and y_i be the actual data for the population and annual number of homicides of city i, $i=\{1,\ldots,S\}$ with S being the total number of cities considered. We assume that the average number of homicides is given by the function

$$\mathbb{E}[Y|N=n_i]\equiv y(n_i)=y_0\,n_i^{\beta}. \tag{H.1}$$

Under the assumption expressed by equation (H.1), the hypothesis to test is that

$$P(Y=y_i|N=n_i)=\mathrm{e}^{-y(n_i)}\frac{y(n_i)^{y_i}}{y_i!}. \tag{H.2}$$

Supposing that the homicides between cities are independent, the maximum-likelihood estimation of the parameters y_0 and β is calculated by computing the likelihood of the data (here denoted as $\{n\}$ and $\{y\}$), and finding the values that maximize it:

$$\begin{aligned} L(\{n\},\{y\},y_0,\beta) &= \prod_{i=1}^{S} \mathrm{e}^{-y(n_i)} \frac{y(n_i)^{y_i}}{y_i!} \qquad \text{(H.3)} \\ \Rightarrow \quad \left.\frac{\partial \ln L}{\partial y_0}\right|_{\widehat{y_0}} &= 0 \\ \left.\frac{\partial \ln L}{\partial \beta}\right|_{\widehat{\beta}} &= 0. \end{aligned}$$

This yields

$$\sum_{i=1}^{S} \left(y_i - \widehat{y_0}\, n_i^{\widehat{\beta}} \right) = 0 \qquad \text{(H.4)}$$

$$\sum_{i=1}^{S} \ln n_i \left(y_i - \widehat{y_0}\, n_i^{\widehat{\beta}} \right) = 0. \qquad \text{(H.5)}$$

We can use appendix H.1 to replace $\widehat{y_0}$ in appendix H.1 in order to have a function of one variable for which we have to find its root $f(\beta) = 0$. To this end, we use the Newton-Raphson method, which consists of finding the root x_{i+1} of the tangent line of $f(x)$ at a given point x_i, and iterating this several times. In mathematical terms, we iterate in our case

$$\beta_{i+1} = \beta_i - \frac{f(\beta_i)}{f'(\beta_i)}$$
$$\beta_0 = seed,$$

where we choose the *seed* to be reasonably close to the real root, for the sequence to converge correctly ($\beta_i \to \widehat{\beta}$).

The procedure to test our hypothesis expressed in equation (H.2) will be the following:

1. Estimate the MLE parameters $\widehat{y_0}$ and $\widehat{\beta}$ from the actual data.
2. Compute the log-likelihood $\mathscr{L}^{(\text{real})} = \ln L\left(\{n\},\{y\},\widehat{y_0},\widehat{\beta}\right)$ of the actual data.
3. Generate R synthetic sample sets $\{y^{(\text{synthetic})}\}$ (each of length S) from the probability distribution of equation (H.2) given the real populations $\{n\}$.
   (a) Estimate the MLE parameters $\widehat{y_0}^{(\text{synthetic})}$ and $\widehat{\beta}^{(\text{synthetic})}$ from the synthetic sample.
   (b) Compute the log-likelihood $\mathscr{L}^{(\text{synthetic})} = \ln L\left(\{n\},\{y^{(\text{synthetic})}\},\widehat{y_0}^{(\text{synthetic})},\widehat{\beta}^{(\text{synthetic})}\right)$ of the synthetic sample.
4. Count the fraction p of the synthetic samples that had $\mathscr{L}^{(\text{synthetic})} \leq \mathscr{L}^{(\text{real})}$. We should understand p as a goodness-of-fit statistic for the Poisson distribution given the scaling relation in equation (H.1).

Country	β	y_0	$\mathscr{L}^{(\text{real})}$	$\max(\mathscr{L}^{(\text{synthetic})})$	$\min(\mathscr{L}^{(\text{synthetic})})$	$\boldsymbol{p}$
Colombia	1.326	1.081e-05	-4485.8	-1562.5	-1686.1	**0.0***
Mexico	0.965	13.922e-05	-4193.8	-2645.7	-2859.9	**0.0***
Brazil	1.213	1.729e-05	-11977.9	-7258.1	-7565.4	**0.0***

Table H.1: Table Showing the Maximum-Likelihood Estimations of the Data Assuming a Poisson Distribution for the Distribution of Homicides Given a City Size. $R = 1000$ Synthetic Data Were Generated. According to These Results, the Plausibility that the Real Data Holds to a Poisson Distribution, $p = 0.0^*$ Is Very Low.

H.2 Results

The maximum-likelihood estimators were calculated with $R = 1000$ synthetic samples according to the mentioned procedure, using a seed of $\beta_0 = 1.0$ in each case, and the results are shown in table H.1. As can be observed, the Poisson distribution can be rejected with high confidence given that $p = 0.0^*$, and we can see that the likelihood of the actual data is much smaller than the smallest of the likelihoods of the synthetic data.

To get a sense of the difference between the real data and the synthetic data, a comparison is shown in fig. H.1. One of the reasons why a Poisson distribution does not describe the data from the figure is the fact that the variance is much greater in the data than in the simulations, which was already evident from fig. 3.4, where the variance (bottom row) did not show a dependence on y_i, in opposition to a Poisson distributed random variable whose mean is equal to its variance.

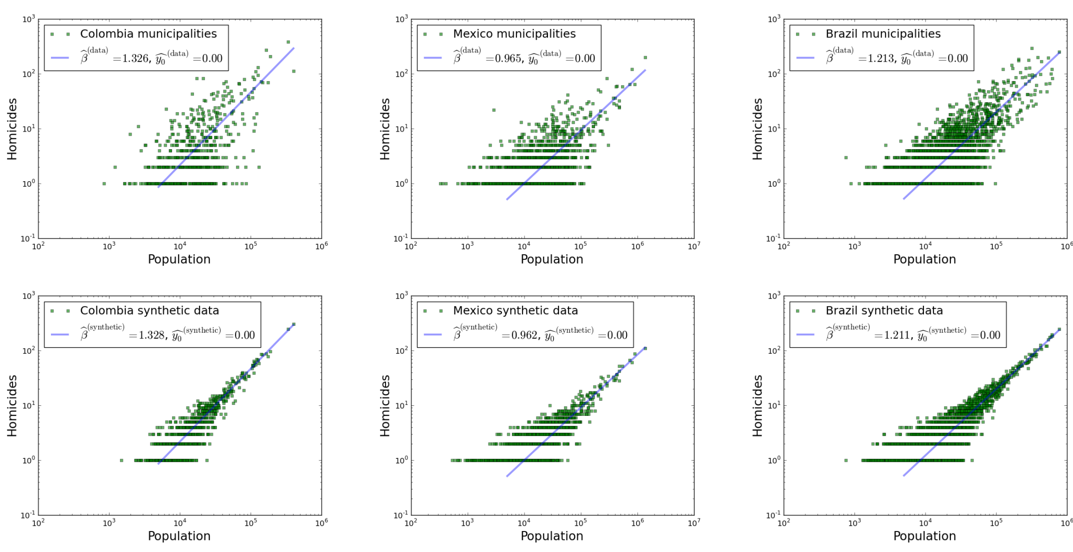


Figure H.1: Top Row: Scatter Log-Log Plots of the Actual Homicides Versus Population in 2007. In All Three Countries Large Deviations from the Fit Are Displayed. Bottom Row: Scatter Log-Log Plots of Typical Synthetic Homicides Versus Population Assuming a Poisson Distribution. It is Clear that the Real Data Shows Greater Variance and Is Not Consistent with a Poisson Distribution. It is Important to Notice that Because of the Logarithmic Scales, the Municipalities with Zero Homicides Are Not Shown.

As a marginal comment, note that the stance here is from a *frequentist* perspective; i.e. we assume one particular possible model of reality, we estimate its parameters which we assume to exist objectively, and generate several random samples from it. A *Bayesian* approach would have assumed that our uncertainty about the model and the parameters should have been expressed through probability priors, and the methodology would have included those considerations.

The results, however, are straightforward in showing that the simple Poisson distribution is not a good model for the data we see in reality.

H.3 Discussion

The method presented here to assess whether a particular distribution describes well the data has proven useful to reject the hypothesis that the annual number of homicides in a city is a Poisson random variable, assuming that on average homicides scale with the population size in a non-linear way.

The greater variance not accounted for by a Poisson distribution (fig. H.1) tells us something about the mechanisms underlying urban homicides. Recently, researchers Sah (1991); Glaeser *et al.* (1996); Gaviria (2000); Calvó-Armengol and Zenou (2004) have pointed out that the dynamics of crime is strongly dependent on the dynamics of interaction. Moreover, they focus on thinking about crime in terms of the decisions of agents to become criminals dependent on the level of criminality of their local social neighborhood. In this spirit they argue that crime is contagious, which would in turn explain the excess of variance seen in the number of homicides.