Quantitative Methods for the Comparative Analysis of Cities in History

Luís M. A. Bettencourt\textsuperscript{1,2,3*} & José Lobo\textsuperscript{4,5†}

\textsuperscript{1}Mansueto Institute for Urban Innovation, University of Chicago, Chicago, 60637, IL, USA
\textsuperscript{2}Department of Ecology and Evolution, Department of Sociology, University of Chicago, Chicago, 60637, IL, USA
\textsuperscript{3}Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, 87501, NM, USA
\textsuperscript{4}School of Sustainability, Arizona State University, Tempe, 85287, Arizona, USA
\textsuperscript{5}ASU–Santa Fe Institute Center for Biosocial Complex Systems, Arizona State University, Tempe, 85287, Arizona, USA

Comparative studies of cities throughout history are one of the greatest sources of insight into the nature of change in human societies. This paper discusses strategies to anchor these comparisons on well-defined, quantitative and empirical characteristics of cities, derived from theory and observable in the archeological and historical records. We show how quantitative comparisons based on a few simple variables across settlements allow us to analyze how different places and peoples dealt with general problems of any society. These include demographic change, the organization of built spaces, the intensity and size of socioeconomic networks and the processes underlying technological change and economic growth. Because the historical record contains a much more varied and more independent set of experiences

\*E-mail: bettencourt@uchicago.edu
\†E-mail: jose.lobo@asu.edu
than contemporary urbanization, it has a unique power of illuminating present puzzles of human development and testing emergent urban theory.

**Keywords:** Urbanization, scaling, economic growth, data, Zipf’s law, Instability, Resilience

1 Introduction

Cities have always held a special fascination to any scholar of human societies. Coincident with the advent of the first cities, we observe the appearance of many technologies and adaptations that, in different forms, are still with us today (1). Thus, the experience of living in cities (2,3) provides a general conducting line throughout history, connecting common phenomena across different societies and thus also identifying features that are truly contextual.

Performing comparative analyses of different societies is always an exercise fraught with challenges. There is the empirical challenge of identifying cultural, social, political and economic traits which can be measured in very different settings. But there is another difficulty when doing comparative analysis which habitually goes unnoticed. The identification of common traits is often conditioned on performance measures, such as rates of economic growth or energy use per capita, which convey a sense of what today we find important (4). Assessing the nature, and even the quality, of ancient societies can easily be biased by using the socioeconomic experience of today’s high-income nations and their recent history. Is evidence for improvements in diet or material conditions in ancient societies to be disregarded because these same societies did not experience high (by today’s standards) output growth rates? Such a stance crudely disregards
many of the extraordinary adaptations and inventions - social, cultural and technological - of earlier societies. Alternatively, conditioning on environmental stewardship and sustainability leads to the opposite conclusion, ranking smaller scale societies that had less impact on their immediate natural environments as having higher quality than most recent societies.

There is however an alternative to such approaches, which starts with much more basic but also more pervasive features of any settled human society \(^{(5,6)}\). A number of recent new ideas, supported by extensive empirical analyses, point to certain quantitative comparisons of basic general quantities that may shed light on a number of key puzzles about the organization, sociality and capacity for adaptation of past urban societies \(^{(5-14)}\). Such puzzles include the relative size, structure and flows between settlements in the same polity, the nature of socioeconomic networks in cities, the spatial organization of settlements, and the nature of change and adaptation in these systems, including processes of economic growth. (Economic growth is here understood to be simply an increase, from one period to the next, in a society’s material output.) What is most important to capture through such comparisons, in our view, is how different societies deal with general problems affecting them all, including energy and resource extraction, and the organization of their socioeconomic networks over space and time \(^{(5,6)}\).

As we look back at history from a modern perspective, shaped by an urban planet with large human population and fast economic growth and technological change, these puzzles become especially poignant: Are pre-industrial societies fundamentally different in the way people lived and interacted? Or are these differences primarily connected to issues of scope, scale and technology?
Can we identify, clearly and empirically, lines of continuity and divergence in the structure and dynamics of urbanizing societies?

These puzzles cannot be answered simply by using the present as the baseline for comparison: what is needed is a framework that makes comparison between the experiences of the past and life in the present conceptually coherent and empirically consistent (5, 6). Here, we explore three strategies for quantitative analysis of settlements throughout history. We discuss how these are undertaken methodologically and their promise for generating a more integrated understanding of our social history as well as an appreciation of each society in its own context. A comparison of the past and the present which is truly based on fundamental processes and features makes it intelligible to use the past and present to discern what the future might be like.

2 Results

Because we are asking for quantitative ways to perform comparative analysis of cities in history we need to obtain data that are consistent across places and times. This remains a challenge, not only because empirical evidence in the archeological record is sparse and mostly associated with durable materials, but also because methods and definitions have naturally varied between many different communities, each dedicated to different periods, using different methods of analysis, etc (15).

Thus, to go forward and attempt any reasonable synthesis, simplicity and clarity are paramount. Simple quantities such as the area of a settlement, its putative population count (based on indepen-
dent measures, such as room counts or amounts of debris), and perhaps other basic quantities related to public spaces or monument construction are usually available through the material record, and have now been measured in several instances (5–9). The analytical advantage of these quantities is that they are reasonably objective and salient features of any human settlement, while leaving plenty of room for varying cultural, political and economic features of different societies (3,4).

For simplicity then, we ask below what we can be inferred from fairly sparse data records, where only a few variables (one, two, ...) are available for each site. This approach also allows us to connect to well-known traditions in demography and geography (16–18), before we attempt to take longer steps towards the end of the paper.

One variable: Demography and the city-size distribution Perhaps the most established way to characterize an urban system quantitatively is by analyzing the statistics of settlement sizes, or equivalently testing the "Rank-size" rule (16–19). This is the simplest of all tests of any quantitative expectation for cities. It requires data on only a single variable, such as the population of each settlement. For this reason, studies constructing the settlement size distribution for many societies are numerous and have been undertaken for decades (16,17,20,21). In many archeological applications, population is replaced by more directly observable proxies, such as the settlement’s area.

The simplest expectation for the rank-size rule (also known as Zipf’s law (17, 22) states that, when cities are rank ordered from largest (rank= 1) to smallest (rank = number of cities in the
system), the size of each city is simply inversely proportional to its rank:

\[ size(rank) = \frac{size_{\text{max}}}{rank^z}, \]  

(1)

where size\(_{\text{max}}\) is the size of the largest city and \(z\) is an exponent. The standard Rank-size rule applies for \(z=1\). This is equivalent to the probability distribution of city sizes taking the form

\[ P(size) = \frac{P_0}{size^{1+z}}, \]  

(2)

where \(P_0\) is a normalizing constant, so that the probability integrates to unity.

Much has been made of the shape of the city size distribution and its meaning. The common exercise deals with the estimation of the rank-size exponent, \(z\), and observing its deviations away from unity. The existence of a distribution of city sizes has been attributed to the (neutral) trade-offs between the benefits and disadvantages accruing from populations agglomerating (19), between economies of scale and costs of movement (18), and a stochastic”preferential attachment” growth process (23). Others have shown that, in some circumstances, Zipf’s law is not a good description of data at all, and distributions in the lognormal family, in particular, may fit the data better (24). This kind of problem is clearly visible in Figure 1, and is discussed below.

Often patterns of settlement size are called *primate* if a single city is much larger than all the others and larger than what the rank size rule would predict. This has often been taken to signal political and economic centralization, in some cases beyond the territory of the settlement system as in the case of empires (25). Primate (or *macrocephalous*) settlement systems of this kind seem to apply to many cases in history, from the Aztecs to contemporary France or England.
Likewise, in many other situations there are several large cities of roughly about the same size (perhaps the Maya, contemporary Spain, Italy, or even Germany (26)). This is sometimes interpreted as a sign of a not fully integrated political or economic system across settlements, with several large cities competing for the "highest" functions associated with the urban hierarchy, such as the central place of government or the dominant (financial) market (27).

Another pattern is a deficit of small settlements relative to what the rank-size rule would predict. This is a common occurrence for most contemporary, highly urbanized settlement systems. To appreciate this consider that, for a system with a largest city of a million (like Rome under Augustus 12), the rank size rule predicts 10,000 towns of 100 people and 1,000,000 with one person. For contemporary systems, with the largest cities in the region of 20-30 million, this would predict way too many small settlements, which are demonstrably not there. This means in practice that the rank-size rule cannot apply across the entire set of settlement sizes, especially for very small ones. For intermediate settlement sizes, some quantitative geographers would attribute this deficit to issues related to the definition of small settlements, many of which they would separate from the orbit of larger places. Varying spatial definitions of cities, usually through different criteria of spatial clustering, can indeed obtain more "Zipfian" city size distributions. This in turn raises issues for of settlement definition, especially for large cities, which are often surrounded by many commuting towns, giving rise to integrated labor markets known in modern settings as metropolitan areas (28).

Despite these interesting interpretations, there seems to be no strong connection between the
relative size of settlements in an urban system and its overall performance, for example in terms of rates of economic or demographic change (29).

The sure lesson that can be derived from the observation of the relative sizes of settlements is very simple. Mechanically, the size of each city measures simply the integrated growth (including periods of decline) over its history, which is essentially a measure of its demographic average growth rate over a long period of time. The simplest rank-size rule states, from this perspective, that all settlements grow at the same rate (20) (if they were created at the same time), another approximate statistical regularity known as Gibrat’s law. Note that this does not have to mean that demographic growth rates are the same for all cities at all time, but simply that over a long period of time these rates converge to the same number, presumably as the result of balances between births, deaths and migration between these towns and cities.

Thus, by comparing the relative size of different settlements, historians and archeologists should be asking whether these were part of the same “demographic” system, connected by mutual migration flows and other networks of exchange and trade. If so, observing something close to the settlement size distribution predicted by the rank-size rule would imply the same average population growth rates for all places, big and small. Then, for example, if mortality rates were higher in larger cities, this would imply a correspondingly larger rate of immigration from smaller places to larger cities (30,31). If it is possible to measure the size distribution of the same settlements at two or more times, then we can moreover compare their relative growth rates during the intervening periods, giving us an empirical basis to rank their relative (demographic) success.
Some additional issues are worth flagging here. For human settlements, spatial areas are not typically proportional to population sizes. How physical space is used socially can be modulated by cultural and physical infrastructure, as well as by technology \(^{(1,2,5)}\). Furthermore, areas can also be measured in different ways, as the surface within the convex hull of the settlement’s putative boundaries \(A\), or as the actual built up area of buildings, streets and other structures, \(A_n\). We discuss these issues next.

Two variables: Settlement Scaling, density and agglomeration  What was life like in the ancient city of Ur? Or in the great city of Teotihuacan? We will never really know for sure, of course. One of the main objectives of archeological and historical research is to reconstruct what social and economic life might have been like from fragmentary information, much of it about the built environment. This is a very difficult type of inference that requires a testable theory of how properties of social and economic life relate to variations in specific characteristics of the built environment.

Settlement scaling theory attempts to do precisely this \(^{(6-9)}\). Developed originally to explain urban scaling properties in contemporary cities \(^{(5,6,10)}\), its ingredients are very general leading to the exciting prospect of the application of its core ideas to settlements in history. The empirical observations on which it is based, as well as its core models, indicate that several basic social economic and infrastructural properties of settlements are interrelated, and can thus be predicted on the basis of comparative analyses of their built environment and estimates of their population size. Empirically, scaling analysis is also very simple, requiring only pairs of variables for each settlement, and the analysis of a familiar xy plot, Figure 2.
If we have two variables for each settlement we can ask for example, how does their built up area depend on their population size: all we have to do is plot one quantity against the other. The answer tends to be non-linear, but well described by scale invariant functions (power-laws), such as

\[ A(N, t) = a(t)N(t)^\alpha. \]  

This can be made linear by a simple transformation to logarithmic variables, or a loglog xy plot, Figure 2.

Theoretical considerations derive the values for the prefactor \( a(t) \) and the exponent for area as \( 2/3 \leq \alpha \leq 5/6 \), depending on the type of settlement and how area is measured (5,6). These expectations are confirmed by empirical analysis of many settlement systems, including in the pre-Columbian Basin of Mexico (7,11), classical Rome (12), Medieval Europe (13), and of course contemporary urban areas (5,10).

The same theoretical framework predicts scaling relations and exponent numerical values for many other quantities, see e.g. Figure 2, including the number of socioeconomic interactions in a settlement, its division of labor, its rate of socioeconomic production and many detailed characteristics of the built environment, such as street length and width and associated transportation costs (5).

In this way, a very straightforward two-variable scaling analysis can reveal commonalities of settlements as socioeconomic networks self-consistently embedded in built spaces. An expansion
of this type of analysis to other settlement systems promises to reveal common quantitative patterns of basic settlement organization and socioeconomic capacity in societies through space and time.

It is also from the perspective provided by these observations and associated theoretical frameworks that we may appreciate any exceptions. For example, an interesting set of questions has been raised by Fletcher about "low density urbanism" (14), specifically in the context of Mayan settlements and Angkor Wat, which appear to show an expansion of their area with population with an exponent, $\alpha > 1$. Therefore such settlements would become less dense the larger they are, not realizing agglomeration effects typical of other cities. A similar, but perhaps more expected pattern also applies to mobile hunter-gatherer camps, but with greater variability. These patterns also vary in time in specific ways, to which we now turn.

**Technological and Economic Change** In modern societies, cities have been a necessary condition for economic growth 32. We say *necessary* because the existence and expansion of cities is not always sufficient for income growth at the national level: there are many episodes, some shorter and some longer, of urbanizing societies experiencing no economic growth. Nevertheless, the association between higher levels of urbanization and larger GDP per capita is one of the strongest empirical results in studies of economic change and international development.

Much work has been done to try to elucidate this connection and better understand the mechanisms of technological change and economic growth generated by urban environments (32,33). However, if the judgment of success is predicated on creating quantitatively precise growth rates, it remains fair to say that the problem is not yet well understood.
Many studies in economic history have also shed light on the circumstances that led to sustained economic growth after the industrial revolution, calling our attention to macroeconomic factors such as the availability of energy on a large scale, political and economic institutions, and the advent of modern science. The study of socieconomic development in the past has also highlighted the role of urbanization \(^{34, 35, 36, 37, 38}\) as have historical experiences of urbanization without growth\(^{39}\). As useful as detailed case studies and historical examinations are, comparative analyses have been hampered partly because of a perceived lack of common empirical evidence within regions and across eras and geographies.

There are, however, a set of facts that may be useful for framing the study of urbanization’s role across time: i) sustained economic growth is a system’s level property (see Figure 2 for China); ii) growth volatility reduces rates of economic growth; iii) very small rates of economic growth are not perceptible over a human lifetime; and iv) the accumulation of material wealth resulting from low-levels of growth are vulnerable to exogenous shocks (such as disease or changes in climate). As a consequence, growth can go unnoticed and remain accidental.

The first point may not be obvious as we often think of rich and poor settlements, even within the same nation or polity. It is, however, generally true that the type of sustained and fast economic growth observed in modern settlement systems is a system level property (so that information, ideas, resources and individuals can flow among settlements), with all cities experiencing about the same annual rate of growth, see Figure 2.

The happy consequence of this observation is that studying systemic economic growth in his-
tory may require only a number of point assessments, which should agree in magnitude whether they were measured in small towns or larger cities. This also means that golden ages often associated with large cities, such as classical Athens or Rome, whether triggered by a technological innovation or by conquest and theft, may not be sustainable unless they induce economic growth across their settlement systems (38). This means, for example, that we should see the living experience of primary producers living in small settlements change so as to enjoy some of the products of large cities and vice-versa in a virtuous cycle of exchange and common development. We know of course that prior to the industrial revolution such periods, if they existed at all, were not associated with large growth rates, and were typically localized in space and time.

The second and third properties of economic growth follow from its character as a stochastic (fluctuating) process. This is a very general feature of collective dynamics of growth, from population biology to financial markets (40). Without going into detailed models for those contexts, quantities such as the resources available to a society (‘wealth’) are expected to grow approximately as

$$\frac{dr(t)}{dt} = (\eta + \epsilon)r(t),$$

where $\eta$, $\epsilon$ are the average growth rate (an approximate constant in time, say 1% a year) and the corresponding stochastic variations, respectively.

Writing the variance of $\epsilon$ as $\sigma^2$ (also known as the square volatility) allows us to integrate the
equation in time to give

\[
\ln \frac{r(t)}{r(0)} = \left( \eta - \frac{\sigma^2}{2} \right) t + \sigma \sqrt{t} \xi(t),
\]

(5)

where \( \xi(t) \) is an approximately normal variable with zero mean and unit variance. The actual growth rate \( \eta - \frac{\sigma^2}{2} \) that results is the geometric mean (not the arithmetic mean!) of growth rates, as is well-known in population biology. This is reduced from the average growth rate by a term proportional to the square volatility, \( \sigma^2 \), that is half the variance of the growth rate, due to its fluctuations over time. Thus, high variability can render any small growth rate zero or even negative, see Figure 3. This means that innovations to reduce instability in the economy are, in the beginning, almost as important as having a positive growth rate in the first place.

With all that said, the final argument we wish to emphasize here is that the growth rate for any preindustrial economy over any extended time period (say decades) was likely very small. Figure 3, based on lead emissions, suggests a value of about 0.17%, certainly lower than 0.3% a year. This translates at the most into a doubling time for the economy of 240 years. This time scale is too long to be felt by anyone - on average at least - in their own lifetimes. Thus, even if slow economic growth was present in preindustrial societies, it was likely too slow for its society to become conscious of it and take measures that could sustain it. The perception would then be one of effectively zero growth, where any positive period would be quickly reversed by fluctuations.

Even if the change in material output of societies in the past had been exponential in nature, the accumulation of wealth could have been greatly set back by disease, climate change or war. And even if the underlying social processes by which agglomerated populations learn, innovate,
and become more productive are the same across eras, societies abilities to deal with the plague or sharp reduction in rainfall are importantly determined by science and technology.

In conclusion, processes of human development and economic growth recognizable to us today were probably at play throughout history, and certainly in most urban societies. However, even in the best of times such rates of change may have been too local, too volatile and too short-lived to be acted upon and sustained, intentionally, over the long term. The search for some of the tell-tale signs of these episodes, especially in the systemic change in living conditions across settlement sizes and ways of living may give us precious new insights into the actual time dependent variability of these effects, and on the human experience in cities during long periods of very slow growth and even decay.

3 Discussion

The history of cities presents us with a bewildering variety of social, economic, political and cultural ways in which human settlements can exist. Making sense of this variation, while at the same time extracting what may be essential across time and space, is a necessary goal not only for a “science of cities” but for a science of human sociality. We have shown how an approach to comparative analysis based on common but determinant variables for human settlements - including population, area and measures of the built environment- has the power to support an analytical narrative relating the earliest settlements in history to contemporary cities and presumably their future forms.
The ambition to develop a theoretical and empirical basis for the study of human settlements through history may invoke in the reader common criticisms of any cross-cultural comparative analysis, specifically that some societies will be judged to be better or worse, and that contemporary high-income societies along with their economic and political systems will be used as standards for evaluation. None of this follows from the strategy proposed here, except the fairly mechanical features that some societies are larger than others - in both their creative and destructive capacity - and so must possess knowledge of, and access to, different types of resources, in ways that are sometimes sustainable and sometimes exploitative. Only by learning formally about this variation can we come to appreciate the range of the human experience in cities.

The approach proposed here then simply connects social and cultural life to some of its most basic material underpinnings, common to all societies in all places. This includes the fact that people exist in space and that their interactions must be structured over space and time in ways that must be compatible with their collective socioeconomic capacity. Evidence from historical and archeological sources have the singular potential to illuminate these issues in ways that contemporary evidence cannot.


Acknowledgments

We are thankful to the speakers and participants of two workshops, *Urbanization without Growth in the Ancient World?* (at the British Academy in Rome, July 2017) and at the Roman Archeology Conference (RAC/TRAC) 2018, where some of the material presented here was originally developed. This work is partially supported by OPEN-AIRE funding linked to the Marie Curie Project Past-people-nets 628818, conducted by Francesca Fulminante (2014-2016).

Author biography

**Luís M. A. Bettencourt** is the Pritzker Director of the Mansueto Institute for Urban Innovation at the University of Chicago. He is a Professor of Ecology and Evolution and the College at the University of Chicago and a Associate Faculty in Sociology. He is also External Faculty at the Santa Fe Institute. His work aims to create novel aspects of urban theory and methods, motivated by
comparative studies of urbanism and the availability of empirical evidence about cities worldwide and throughout history.

José Lobo is a Senior Sustainability Scientist at the Julie Ann Wrigley Global Institute of Sustainability, an Associate Research Professor at the School of Sustainability and a Faculty Associate, Department of Economics, W. P. Carey School of Business. He is interested in determinants of metropolitan economic performance and location-specific economic growth; the application of machine learning, data mining and spatial statistics methods to the study of socioeconomic data; causes and consequences of urban size and scale; and how the characteristics of individuals, organizations, institutions and social networks interact to create "regions of innovation."
Figure 1: The city size distribution of Metropolitan Areas in the USA in 2010. A. histogram of city sizes (red dashed line is lognormal distribution fit); B. Rank-size rule (black line is Equation (1)). We see that in the USA the rank size rule approximately describes the relative distribution of large cities but fails to account for an overabundance and then deficit of progressively smaller towns.
Figure 2: Economic and demographic growth in modern cities, such as those of China (shown) is a property of the urban system. Panel A shows about 20 years of data for Chinese Prefectural cities (colored dots), and the scaling of GDP with population size (solid line shows the scaling relation). Panel B shows the same scaling after system wide growth (yellows squares) is subtracted. This growth is shown in panel C for both GDP (red) and Population (yellow) and versus each other on panel D.
Figure 3: Production and volatility measured by lead emissions (measured in Greenland ice cores, Ref. \((41)\)). Panel A shows estimated emissions over a long historical period. Panel B shows the corresponding growth rate in emissions (orange is a running average). Red vertical lines delimit the period between 150BC and 150AD, associated with a rise of the Roman Empire. For this period the effective growth rate is very small due to high volatility. The annual average growth rate is about \(\eta \approx 0.17\%\).