Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The Rise & Fall of Urban Concentration in Britain: Zipf, Gibrat and Gini across two centuries

Ronan Lyons and Elisa Maria Tirindelli

TEP Working Paper No. 0522

September 2022

Trinity Economics Papers
Department of Economics

The Rise & Fall of Urban Concentration in Britain: Zipf, Gibrat and Gini across two centuries

September 13, 2022

Abstract

While city size and growth are the subject of a substantial literature, consensus is lacking on the extent to which Zipf's Law or Gibrat's Law holds across space and time. We examine city size, rank and growth in Britain 1801-2011 and show conclusions depend on city definition, sample cutoff and regression methods. We find Zipf's Law cannot be rejected under the strongest combination of data and methods, unlike if other data or methods are used. Across Zipf, Gibrat and Gini analyses, we find that urban concentration in Britain peaked in the mid-19th century but fell 1861-1911 and 1951-1991.

Keywords: Great Britain; Zipf's Law; urban growth; Gibrat's Law.
JEL codes: N9; O18; R11; R12.

1 Introduction

Over the last two centuries, cities have become integral to the global economy and to human experience. The fraction of the global population living in cities is estimated to have been just 10% in 1800 but, by 2010, more than half the world's population lived in cities (UN-DESA, 2018). Understanding patterns of city growth, therefore, will be central in accommodating an additional three billion city-dwellers over coming decades. At the heart of patterns of city size and growth are Zipf's Law (Zipf, 1965) and Gibrat's Law (Gibrat, 1931). Zipf's Law holds that the relationship between a city's size and its rank is unit elastic – in other words, a 10% increase in a city's population leads to a 10% fall in its rank. It is described by Krugman (1996) as "one of the most overwhelming empirical regularities in economics". Gibrat's Law holds that there is no systematic relationship between a city's initial size and its subsequent growth rate.

In this paper, we examine the relationship between city size and both city rank and subsequent city growth, over the last two centuries in (Great) Britain. We supplement these investigations of both Zipf's and Gibrat's Law in the world's first urbanized economy with an analysis of urban concentration using the Gini coefficient, as suggested by Henderson and Wang (2007). To do this, we assemble for the first time comprehensive data for all cities in Britain across 19 Censuses between 1801 and 2011. Our dataset allows us to measure cities in four different ways – local government districts, unitary authorities, primary urban areas, and travel-to-work areas – across four different statistical cutoffs for city size. In addition, we also compare across methods suggested by the literature, including parametric and non-parametric methods, and in the case of the "Zipf" exponent, with and without correcting for bias in the coefficient and also in the standard error.

Both Zipf's and Gibrat's Laws are the subject of intense debate among researchers, with deep implications for policymakers. Our paper contributes to this extensive literature on city size, rank and performance and in particular aims to bring clarity to researchers looking to test for the presence of either law in empirical settings. Seminal contributions include Auerbach (1913), who identified a size-rank relationship for cities, Gibrat (1931), who showed that a proportionate growth process delivers a Pareto distribution in the upper tail, and Zipf (1965), who found a similar relationship more broadly, including in word frequency.[1] Gabaix (1999a) connected Zipf's Law and Gibrat's Law, by noting that the random growth in city populations (or population shares) described by Gibrat would lead to the city size-rank relationship described by Zipf. While many authors note the occurrence of Zipf's Law in very different geographies and time periods, some have emphasised deviations. Using the well-known Bairoch et al. (1988) dataset of European city populations over time, Dittmar (2020) shows the emergence of Zipf's Law across Western and Eastern Europe at different points in the transition from the medieval period to the modern period. Soo (2005) uses size-rank distributions at a country level to build a dataset of coefficients and concludes, based principally on municipal definitions of cities and OLS estimators, that the coefficient on the Pareto distribution differs from one in the majority of cases.

However, Soo's results for urban agglomerations – available for a much smaller number of countries – suggest the opposite result, highlighting the importance of city definition and methodology. Dingel et al. (2019) examine the city size-rank relationship for three large developing countries – Brazil, China and India – using nightlights to es-

[1] For a review of the literature that started with Gibrat's work, see Sutton (1997).

timate true urban agglomerations. They find that, while the size and rank of official municipalities does not follow a linear relationship, the relationship measured using urban agglomerations does. The literature on Gibrat's Law is equally diverse. An important contribution to this literature is Glaeser et al. (2014), who explore the dynamics of county growth in parts of the USA 1790-2000. While they find evidence in favor of Gibrat's Law for the sample as a whole, this does not hold for long sub-periods, with less populous counties growing faster before 1860 and after 1970. This raises the possibility that occurrences of Gibrat's Law are an artefact of the accidental balancing of centripetal and centrifugal forces over different time periods.

We believe our paper makes five principal contributions. Firstly, ours is the first study in the literature to examine Zipf's and Gibrat's Laws in Britain over the long-run. Britain is of wider interest as it was home to the Industrial Revolution and, related, the world's urbanized economy. Secondly, we highlight the importance of research design in examining whether Zipf's Law holds, with spurious results common where arbitrary city definitions or cut-offs are used, in addition to bias and false precision. Using the same setting, we show the Pareto exponent can vary dramatically across 48 specifications in total: three different definitions of city available in our setting, four different sample cut-offs, and with four sets of results for each of these twelve unit-cutoff pairings – the standard (biased) OLS estimator and the unbiased estimator, each with unadjusted and adjusted standard errors (SEs). Thirdly, we find that we are unable to reject Zipf's Law in any of the 19 Census years under the strongest set-up – using primary urban areas, with a conservative cutoff, the unbiased estimator, and adjusted SEs. This is due not only to the point estimate but also to the necessary lack of precision in an urban system with approximately 60 true cities. Fourthly, our analysis of whether Gibrat's Law holds finds evidence against it holding in the period 1861-1911, where there is a strong negative relationship between initial size and fifty-year growth. In later periods, there is only modest evidence for such a link. Finally, across all three sets of analysis, we find evidence of rising inequality in city populations in the early 19th century, followed by greater compression across cities in the following century and a half. This is seen both in the point estimate of the Pareto exponent, which falls until 1861 then rises, and in the Gini coefficient of urban population, following Henderson and Wang (2007). Similarly, deviations from Gibrat's Law point to smaller cities enjoying faster growth after 1861 (and in particular before 1911 and after 1951).

Our paper is structured as follows. In Section 2, we review the underlying theory, both of power laws more generally and relating to exponents and cutoffs for samples, more specifically, and briefly review the existing empirical literature on the city-size distribution, paying particular attention to the city definitions, sample cutoffs and estimation methods used. Thereafter, in Section 3, we introduce our data on British city populations since 1801, and in Section 4, we outline the results of our empirical analysis, before the paper concludes.

2 Theory & Evidence

2.1 Power Laws

Power laws are ubiquitous in nature. Newman (2005) lists a number of physical, biological, and man-made phenomena where power laws are observed, including distributions of size frequency and volume. Examples are found in the size of earthquakes, moon craters, solar flares, computer files, and wars; the frequency of use of words in any human language and of personal names in most cultures; and the volume of papers scientists write, of citations received by papers, of visits to web pages, of sales of books and almost all branded commodities.

Mathematically, a power law function is a relationship between two quantities, where a relative change in one quantity results in a proportional relative change in the other. More precisely, one quantity varies as the power of another, independent of their initial size:

$$f(x) = ax^{-k}$$

The fundamental characteristic of this kind of relationship is scale invariance. This implies that scaling the argument of a power law function causes a proportionate scaling of the function itself:

$$f(cx) = a(cx)^{-k} \quad \text{for } k > 0$$

This is important because it implies that we can observe a linear relation in log-log, which is the most straightforward way to test for the existence of power law behaviour.

Much of the interest in power laws comes from the study of the probability distribution and its application to probability theory and statistics. Strictly, a power law function cannot be a probability distribution as, for any value of the exponent, its integral diverges either in zero or infinity. Consequently, it is necessary to define its support greater than a lower cutoff x_{min} (or smaller than an upper cutoff), and multiply it by a scaling parameter C, so that its integral meets the necessary unity condition of a probability distribution.

A Pareto distribution is a particular type of power law, which is mostly used in social science. If X is a continuous random variable following a Pareto distribution, then its density distribution function is given by:

$$f_X(x) = \frac{\zeta x_{min}^{\zeta}}{x^{\zeta+1}} \mathbb{1}_{[x > x_{min}]}(x)$$

and its cumulative distribution is:

$$P(X < x) = \int_{x_{min}}^{x} \frac{\zeta x_{min}^{\zeta}}{x^{\zeta+1}} dx = 1 - \left(\frac{x_{min}}{x}\right)^{\zeta}$$

Where the Pareto exponent (ζ) takes a value, in absolute terms, of 1, this is called Zipf's law. This regularity is found in a variety of situations and it is so called because it was initially observed by Zipf (1965) in the distribution of words' length.

Our focus is city size. To understand how Zipf's Law can be applied in this context, consider the probability (P) that the size of a given city S is greater than a certain value s. Where G is the survival function of a city's size and r is the city's rank, this probability is proportional to its rank:

$$G(S) = P(S > s) \propto r(S) \tag{1}$$

Where $G(S)$ takes the form of a power law and θ refers to its exponent, then:

$$G(S) = \frac{k}{S^{\theta}} \tag{2}$$

Combining equation (1) and (2) together, it follows that:

$$G(S) = P(S > s) = \frac{k}{S^{\theta}} \propto r(S) \Rightarrow r(S) = \frac{k}{S^{\theta}} \tag{3}$$

In this way, we obtain a specific inverse relationship between rank and the size. Where $\beta = 1/\theta$, $G(R)$ is the distribution of rank and k is a normalising constant, this relationship can also be expressed in the opposite way, as follows:

$$G(R) = P(R > r) = \frac{k^{1/\theta}}{S^{1/\theta}} \propto S(r) \Rightarrow S(r) = \frac{k^{1/\theta}}{S^{1/\theta}} = \frac{k^{\beta}}{S^{\beta}} = \left(\frac{k}{S}\right)^{\beta} \tag{4}$$

This gives the exact same relation as in (3) but reversed, i.e. the probability that the rank R is greater than a given r is proportional to city's size. Zipf's Law is said to hold if $\beta = -1$. An important distinction is needed between rejection of Zipf's Law, where the exponent is statistically different from -1 but the relation may still be log-linear, and rejection of a Pareto distribution, where the log-linear relationship is rejected altogether.[2]

Before Gabaix (1999a), the unit value of the distribution parameter had remained unexplained. Gabaix's random growth model offers a theoretical explanation for the emergence of Zipf's law, both for cities and for other phenomena. He shows that, if we assume one (proportional) growth process for cities above some minimum size, i.e. that cities grow independent of their size (Gibrat's Law), then the steady state distribution of such a process is a power law with exponent 1. The conditions for this to hold are that cities grow on average at the same pace and with the same variance and that the smallest city in the distribution is very small as a proportion of the total urban population. Most other models in more recent papers that involve the Zipf exponent, such as those by Córdoba (2008) and Dittmar (2020), ultimately rely on the same underlying principle to explain its emergence. An exception is Rossi-Hansberg and Wright (2007), who use a general equilibrium model where cities specialise in particular final goods. In their model, Zipf's Law emerges as cities reach efficient size given their specialization, but only where labour is perfectly mobile and not a factor of production.

2.2 Cutoffs and Exponents

As described above, two factors characterize a power law distribution: its exponent and the lower or upper cutoff – in the case of city size distributions, the lower cutoff is the minimum city size. Despite its importance and its relationship with the value of the exponent, there has been very little debate on the estimation of the lower cutoff (Eeckhout, 2004). The majority of the related literature uses a data-driven cutoff (for example: Davis & Weinstein, 2002; Dittmar, 2020; Dobkins & Ioannides, 2001; Dobkins, Ioannides, et al., 2000; Eaton & Eckstein, 1997; Ioannides & Overman, 2003; Soo, 2005). An alternative approach, used by Eeckhout (2004), is to set the cutoff at an arbitrary level to encompass different quantiles of the population. Only Muller (2016), in examining air pollution, uses a statistically defined cutoff.

Given the importance of the cutoff, we set out below four different approaches to its definition. These four approaches are then used to inform the empirical analysis.

1. **Level cutoff**: This is the most common method employed in the literature so far (see for example Dittmar, 2020; Dobkins, Ioannides, et al., 2000; Ioannides & Overman, 2003). Particularly where the empirical setting is the USA, a value

[2] In his paper, for example, Dittmar (2020) fails to reject a Pareto distribution on the Bairoch dataset after 1500 but he does not test the value of the exponent. This implies, strictly speaking, that he is testing for the emergence of a Pareto distribution but not of Zipf's Law, following the glossary provided by Gabaix (2009).

of 50,000 is most common, reflecting the fact that the US Census Bureau reports data on Metropolitan Areas of 50,000 inhabitants and above, from 1898. This level cutoff is then constant throughout the period of analysis. While easy to understand, it means that, where population levels change significantly over the period of analysis, the composition of cities included over time can change significantly, relative to the distribution as a whole.

2. **Fraction cutoff**: An alternative heuristic approach is to take a certain fraction of the distribution in each year; Cheshire (1999) mentions this as one of the options to set the cutoff. This has the advantage over a level cutoff of a relatively even sample size, especially over longer periods of analysis, during which city populations may have changed by an order of magnitude. Another feature of this approach is its focus on the top fraction of the distribution; the relationship between city size and rank may differ within this set of cities compared to the full distribution of cities.

3. **Conservative cutoff**: Clauset et al. (2009) describe a process for selecting the cutoff. Firstly, the power law exponent is estimated for each possible x_{min} in the dataset. Secondly, using a Kolmogorov-Smirnov test to calculate distance, the cutoff is chosen as the one in the exponent-cutoff pair which minimises the distance to a theoretical power law distribution with the same parameters. For each x_{min}, the test also returns the p-value for the null hypothesis of a Pareto distribution. Of the tests, it tends to keep a bigger portion of the sample and is, therefore, referred to as the conservative cutoff.

4. **Deviation cutoff**: The fourth cutoff uses the method outlined in Gabaix and Ibragimov (2011) and applied in Rozenfeld et al. (2011). Here OLS is used to estimate the relationship (in logs) between a city's rank and both its size and the square of size minus γ, where $\gamma = (cov(log^2(size), log(size))/(2var(log(size))$. More formally:

$$log(Rank - 0.5) = \beta_0 - \beta_1 log(size) + \beta_2 log(size - \gamma)^2 \tag{5}$$

   As explained in Rozenfeld et al. (2011), the test formalizes the intuition that high values of β_2 indicate deviations from a power law because, in the limit, a true power law will have $\beta_2 = 0$. They outline a critical value for β_2, at the 1 percent confidence level. If the absolute value of β_2 is greater than this critical value, β_c, the null hypothesis of a power law is rejected. This allows the calculation of a sample size, based on repeatedly extending the sample until the coefficient β_2 is statistically significant.

For a given sample size, the main approaches for estimating the Pareto exponent and the Gibrat coefficient are described below.

Zipf's Law Despite the problems pointed out by Gabaix and Ioannides (2004), the OLS estimator is the most common method used in the literature for the estimation of the Pareto exponent (see, among others, Eeckhout, 2004; Rosen & Resnick, 1980; Soo, 2005, and Table 1, where we summarize 22 of the key papers in the literature). The issues arising with this estimator are its biasedness and its underestimation of the standard error. One can interpret the origin of the bias in the following way: the expected value of the ratio between $S_{(2)}$ and $S_{(1)}$ is 0.5, but the smallest 95% confidence interval for $S_{(1)}/S_{(2)}$ is [1,20]. So typically, the value of $S_{(1)}$ will be above the value predicted by the linear regression with slope -1. In other words, the size of the largest city will look "too big". The underestimation of the standard error on the other hand derives from the fact that the ranking procedure creates positive correlations between the residuals, whereas the OLS standard error assumes that the errors are independent.

An alternative is to use the Hill estimator, which is the maximum likelihood estimator for a power law, available in closed form. Gabaix and Ioannides (2004) show that it also delivers a biased standard error. An easy correction for both issues in the OLS case – bias in the estimator and underestimation of the standard error – is provided in Gabaix and Ibragimov (2011) and Gabaix and Ioannides (2004). Here, we use the OLS estimator with and without the bias and standard error correction proposed by Gabaix, with our preferred combination being the OLS estimator corrected for bias and with corrected standard errors. [3]

Gibrat's Law Concerning investigations of Gibrat's Law, the literature does include examples of parametric estimations, such as linear spline estimation (see Desmet & Rappaport, 2017) and bin-size dummy regression (see Michaels et al., 2012). However, they are less frequently used than non-parametric estimation and less immediate in their interpretation. Therefore we favor more common non-parametric approaches, and in particular the Nadaraya-Watson estimation. Leading papers in this literature use the Nadaraya-Watson estimator, which provides a visually straightforward interpretation of the results (see, for example, Desmet & Rappaport, 2017; Ioannides & Overman, 2003).

2.3 Literature

A vast literature has arisen since Zipf and Auerbach, examining not only the size of the exponent in the power-law distribution of city size and rank but also, following Gabaix, testing whether Gibrat's Law holds, i.e. whether population growth is systematically related to existing city size. It is far beyond the scope of this paper to try to summarize the literature in its entirety. Instead, we focus on both seminal and recent contributions to this literature, paying particular attention to tests of Zipf's and Gibrat's Laws. We report in this section a brief summary of the more extensive literature review given in Appendix A.

We distinguish between short and long-run analysis and, among short-run analyses, between locations. We define the "short-run" literature to be empirical analyses using less than 50 years of data on urban populations and review a range papers that meet these criteria, distinguishing where possible between three sets of analyses, reflecting the choice of urban unit. The first using officially-defined municipalities to delineate cities, the second using metropolitan areas, while the third using definitions of cities built from satellite imagery. The second part of our literature review covers empirical analyses of city growth and size distributions in the long-run, which we define to be over a period of at least half a century. We review these separating into four broad categories based on their (principal) region of analysis: global studies; the Americas; Asia; and Europe.

[3] Alternatives, not covered in this analysis, are estimation with maximum likelihood methods is covered in the appendix and non-parametric approaches, which allow the Zipf exponent to vary by city size. The latter include the local Zipf exponent (Gabaix, 1999b; Ioannides & Overman, 2003) and the Theil estimator, as per Dittmar (2020).

At first glance there appears to be little consistency in the literature. The cases of India and China are emblematic: Schaffar and Dimou (2012) and Chauvin et al. (2017) both find that Zipf's Law does not hold in either country, but they then disagree on whether Gibrat's Law does. Dingel et al. (2019) are not able to reject Zipf's Law in either economy when they use night-lights to define cities and note that they would have if they had used administratively defined cities. Jiang et al. (2015) make a similar finding and raise the importance of scale: Zipf's Law appears to be a weak descriptor of small urban systems but more relevant for larger urban systems. This is intuitive: it seems unreasonable to expect the urban structure of Andorra, home to fewer than 80,000 residents of whom more than half live in the metropolitan area of the largest city, to match precisely the principal characteristics of the urban structures of the continental US or China, vastly larger geographically and economically.

As the examples of India and China suggest, the seeming babel of empirical results relating to Zipf's and Gibrat's Laws stems, at least in part, from differences in the underlying unit being measured, the cutoffs employed on those units, and the empirical methods chosen. As shown in the summary of the short-run literature in the appendix, where larger and less arbitrary urban boundaries are used, it becomes harder to reject Zipf's Law in particular. This is particularly the case in the newer literature using satellite imagery, such as nightlights. Reflecting the literature, as well as economic rather than political realities, our focus is on metropolitan areas, not municipalities. For such units, there is by and large clarity that a power-law relationship exists between rank and size, at least when the appropriate cut-off is employed.

In terms of method for testing Zipf's Law, as outlined above, the state-of-the-art has progressed substantially beyond basic OLS and now includes corrections to both the point estimate and the standard error, as well as non-parametric methods. The latter correction involves use of the sample size itself. When combined with the use of larger urban units, which are by definition fewer in number, one consequence is far less certainty about the parameter estimates. This creates a tension in longer-run analyses between wider confidence intervals, which can mean an inability to reject a parameter estimate of 1 over time, and discerning trends in the point estimate; this is a point we return to in our analysis. For this reason, we add to our analysis trends in the Gini coefficient, as suggested by Henderson and Wang (2007).

The long-run literature typically focuses more on Gibrat's Law than Zipf's. It includes a number of analyses of US urban growth over the long-run, with some evidence in favor of Gibrat's Law but with numerous caveats. Analyses of other urban systems, including Brazil, China and Japan, find support for Zipf's Law but again Gibrat's Law appears to hold only in certain cases. In research that overlaps in setting somewhat with ours, Klein and Leunig (2015) examine urban growth in England 1761-1891 and test Gibrat's Law. They find that it is violated consistently, although this appears to be driven by their choice of unit and lack of any cut-off. Overall, the long-run pattern of urban growth dynamics and the size-rank relationship in Britain, the world's first urbanized industrial economy, remains largely unknown and this is the focus of our analysis.

3 Data

3.1 Administrative Context

To understand the relationship between a city's size and its rank and future growth, we use detailed data on British city populations from 1801 to 2011. The underlying data stem from the UK's 22 decennial censuses. There are at least two key attributes of the British setting that are worth noting. Firstly, because of its island nature, the geographical scope of the larger political unit is fixed.[4] Secondly, Britain was the location of the first Industrial Revolution and, consequently, the first economy to transition to majority-urban. This gives Britain the longest-running panel series for an urbanized economy.

The oldest administrative units in Britain are civil parishes, which in many instances have ancient roots, dating back to feudal times. Civil and religious parishes overlapped until the Poor Law Amendment Act (1866), which defined "civil parishes" to be any unit that levied its own rate, for the purposes of poor relief; this definition included not only ecclesiastical parishes but also other units, including townships. The *Local Government Act* (1888) created larger administrative counties (and county boroughs) as units for local government, which often resembled older traditional (or ceremonial) counties. Under a successor Act in 1894, administrative counties were subdivided, into units known as urban and rural districts. The *Local Government Act* of 1972, together with the *London Government Act* of 1963 and the *Local Government (Scotland) Act* 1973, reformed the make-up of districts in Britain. In England, for example, it split 314 districts into metropolitan (34) and non-metropolitan (244) categories, as well as London boroughs (32) and two other *sui generis* districts.

Under the *Local Government Act* of 1992, unitary authorities were formed in England and Wales, with council areas formed in Scotland, after the 1994 *Local Government etc. (Scotland) Act*. These are local authorities responsible for the provision of all local government services within a (local government) district. This means that, since 1992, a new spatial taxonomy has existed that reflects urbanization within districts: under the 1992 Act, larger towns can have separate local authorities from the less urbanised parts of the same districts. This Act continues to be used, with changes in six ceremonial counties between 2019 and 2021. In Dorset, for example, the number of unitary authorities was reduced from eight to two in 2019. This reduction involved the merger of two existing (urban) unitary authorities and one non-metropolitan district into a new authority (Bournemouth, Christchurch and Poole), while the five remaining non-metropolitan districts were merged to form Dorset Council (UK Houses of Parliament, 2018).

The example of Dorset above highlights the challenges of using municipal boundaries to examine the city size-rank relationship. Adjustments to the units used to officially define cities are not incremental but occur instead in less frequent but more substantive reforms, such as those of 1972, 1992 and (in the case of Dorset) 2019. In order to reliably estimate the size-rank relationship, it is important to use spatial units that are not only consistent over time but also reflect the true extent of urban agglomerations. In the case of the Bournemouth, Christchurch and Poole unitary authority established in 2019, the Christchurch area had previously been a non-metropolitan dis-

[4] Relatively few other countries have fixed geographical boundaries over the same time period. The United Kingdom included the full island of Ireland from 1801 to 1921 and Northern Ireland since. We focus here, however, on Britain, which has been in the same political unit since 1707.

trict. Its inclusion was legally defined in the relevant Statutory Instrument by reference to five "electoral divisions".

Electoral divisions (or wards) are the spatial units used for Britain's administrative geography: all higher administrative units are built up of whole electoral wards or divisions. They are used for parliamentary constituencies and the EU's Nomenclature of Territorial Units for Statistics (NUTS) regions, as well as for the unitary authorities (in England and Wales; council areas in Scotland) and the metropolitan and non-metropolitan districts mentioned above. As of the late 2010s, the UK had over 9,000 electoral divisions/wards, with an average population of roughly 5,500 in each. These form the basis of the *Vision of Britain through Time* database (VOB) we use for our analysis (Southall, 2017). The VOB database brings together historical surveys of Britain, in particular Census Reports, to make information on population by geographical unit publicly available. The database involved conversion of named areas into areas with consistent boundaries over time. To do this, VOB combines accurate information from the 2001 Census with County Administrative Diagrams, published from 1900 and extended back in time using Registrar General maps.

3.2 Spatial Taxonomies

As noted in the literature review, the choice of urban unit may affect the results of tests of Zipf's and Gibrat's Law. For that reason, we use three different geographical units in our analysis, all available from the VOB database and each of which is related to the concept of a city: "Local Government District", "Unitary Authority" and "Primary Urban Area".[5] Summary statistics for all three levels of unit are given in Table 2.

1. **Local Government Districts (LGDs)** are the most granular unit we use in our analysis and also the unit that varies most over time with administrative changes. Data are available on population by LGD for each British census from 1851 to 2011. The number of urban districts rises from 527 in 1851 (with a median population of 5,109) to a peak of 1,143 in 1921 (median population: 8,595). Thereafter the number falls, in particular after the reforms of 1972, when the number of units falls from 895 to 461. There were 347 LGDs in Britain in 2011.

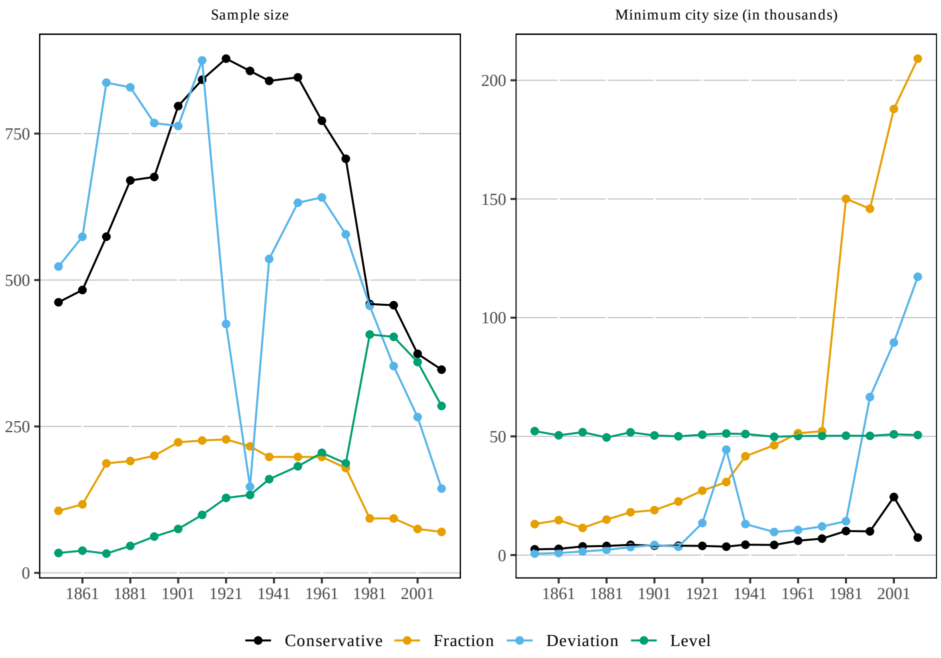
2. **Unitary Authority (UAs)** are, as described above, a larger spatial unit than LGDs and were reported from 2001, the first Census after the reforms of the 1990s. Using population by electoral division across Censuses, the *Vision of Britain* database has calculated the population for each UA by Census year from 1801 to 2011. This means that, unlike LGDs, UAs are fixed spatial units across the entire period of analysis. There are 379 or 380 UAs in most Census years.[6]

3. **Primary Urban Areas (PUAs)** are a spatial unit defined by the Centre for Cities to reflect the "built-up" area of a city (Centre for Cities, 2021). To do this, they aggregate UAs (for England; for Scotland and Wales, in almost all cases, the corresponding local authority area is used). There are 61 PUAs available for almost every Census year from 1801 to 2011. Given their spatial consistency over time, and given their ability to reflect the full extent of an urban agglomeration, these are our preferred unit.

[5] A fourth unit type, "Travel To Work Area", exists in the VOB database, but only for three Census years between 1991 and 2011.

[6] Data for Scottish UAs are missing for 1891.

Figure 1: Local Government District



Note: The left panel of this figure shows the number of Local Government Districts included in the sample for each census year for each cutoff methods. The right panel shows the size of the smallest Local Government Districts included in the sample for each cutoff method.

3.3 Samples by Cut-off

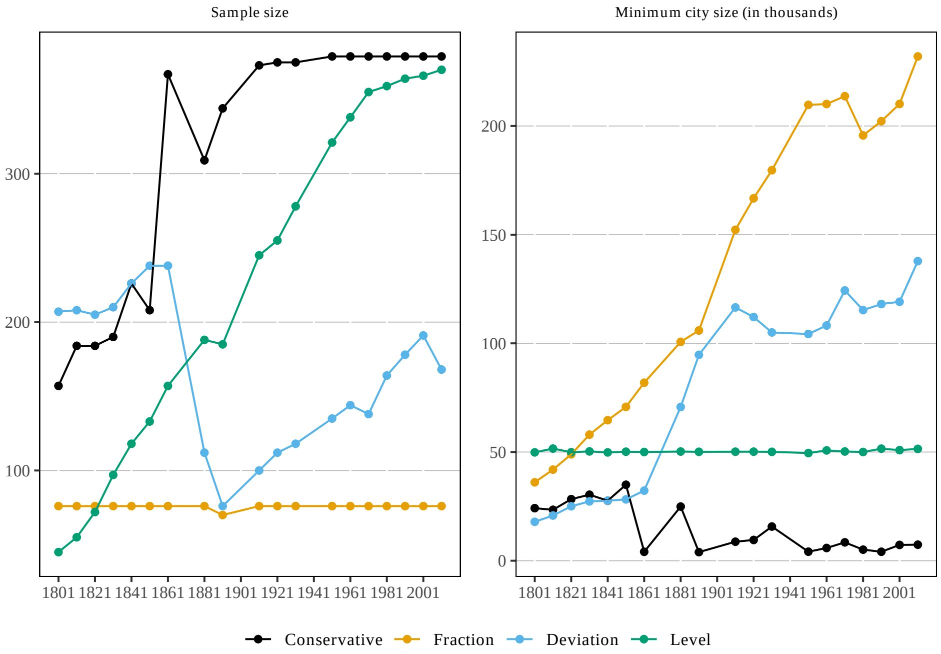
Section 2.2 outlined four possible cutoffs that can be applied to datasets of city populations: what we term the *Level*, *Fraction*, *Conservative* and *Deviation* Cutoffs. Further, Section 3.2 outlined three different spatial definitions of city. This gives twelve different combinations of spatial units and cutoff methods, what we term *unit-method pairs*. Figures 1-3 present for each Census year for which they are available, the number of cities included in the analysis, and the minimum city size, by cutoff, for each of the three levels of spatial units.[7] In each, the black line represents our preferred unit (PUAs).

Using LGDs (Figure 1), the sample size to be included in the analysis is quite variable, both over time (given a cutoff method) and across methods. For example, of 347 LGDs in 2011, all would be included under the *Conservative* cutoff, 285 with a 50,000 *Level* cutoff, 144 under the *Deviation* cutoff, and 70 under the *Fraction* cutoff. Because of their nature, and a combination of the trend in the total number of LGDs and rising populations over time, the *Level* and *Fraction* cutoffs trend in different directions: the *Level* cutoff implies a sample of fewer than 100 before 1921, while the *Fraction* cutoff is at or above 200 for the period 1891-1961. While the minimum city size is 50,000 throughout under the *Level* cutoff, it rises from 13,000 in 1851 to 209,000 in 2011 for the *Fraction* cutoff. Under the *Conservative* cutoff, minimum city size is (roughly) between 3,000 and 4,000 until 1931 and close to 7,000 thereafter. The *Deviation* cutoff has a 100-fold change in minimum city size: from 1,000 in 1851 to 117,000 in 2011.

For UAs (Figure 2), which in our dataset are constant over time with a full sample

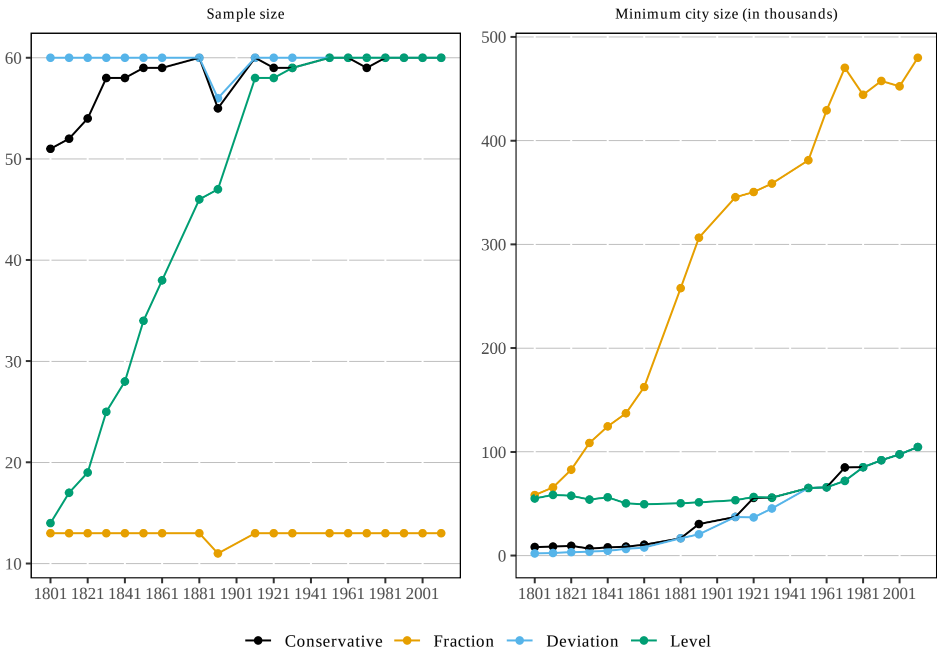
[7]Tables 3-6 present the same information in tabular form in the Appendix.

Figure 2: District/Unitary Authority



Note: The left panel of this figure shows the number of District/Unitary Authority included in the sample for each census year for each cutoff methods. The right panel shows the size of the smallest District/Unitary Authority included in the sample for each cutoff method.

Figure 3: Primary Urban Area



Note: The left panel of this figure shows the number of Primary Urban Area included in the sample for each census year for each cutoff methods. The right panel shows the size of the smallest Primary Urban Area included in the sample for each cutoff method..

size of 380, again under the *Conservative* Cutoff, almost all are included from 1911 on; earlier than this only a smaller share is included (155 in 1801). The minimum city size is quite volatile, at 20,000 or more before 1891 (with one exception) and below 10,000 thereafter (again with one exception). The *Deviation* cutoff yields a sample size of at least 200 before 1881, closer to 100 for the following Censuses, and rising again to close to 200 by the turn of the millennium. Minimum city size increases from less than 30,000 before the 1860s to more than 100,000 from 1911 on. By construction (given the fixed number of spatial units), the *Fraction* cutoff gives a consistent sample size of 76 throughout. There is an order-of-magnitude rise in the minimum city size over the first century, from 18,000 in 1801 to 117,000 by 1911, after which the minimum size is largely stable. Lastly, the *Level* cutoff yields a growing sample over time, from 45 in 1801 to 370 in 2011.

Lastly, there are the 61 PUAs (Figure 3), our preferred spatial unit. The *Conservative* and *Deviation* cutoffs include almost all of these throughout the two centuries, while (by construction) the *Fraction* cutoff only includes 13 each year. The *Level* cutoff trends up over time, starting at 14 and rising to 58 (or higher) from 1911 on.

There are two features that are appealing *a priori* in judging the various cutoffs and spatial units, given the time and geographical setting. Firstly, the best combination of spatial units and cutoffs are likely to exhibit a consistent sample size over time, reflecting the lack of entirely new urban agglomerations in the setting under consideration. Secondly, the strongest unit-method pairs should show a steadily rising minimum city size, corresponding to Britain's growing urban population over the period.

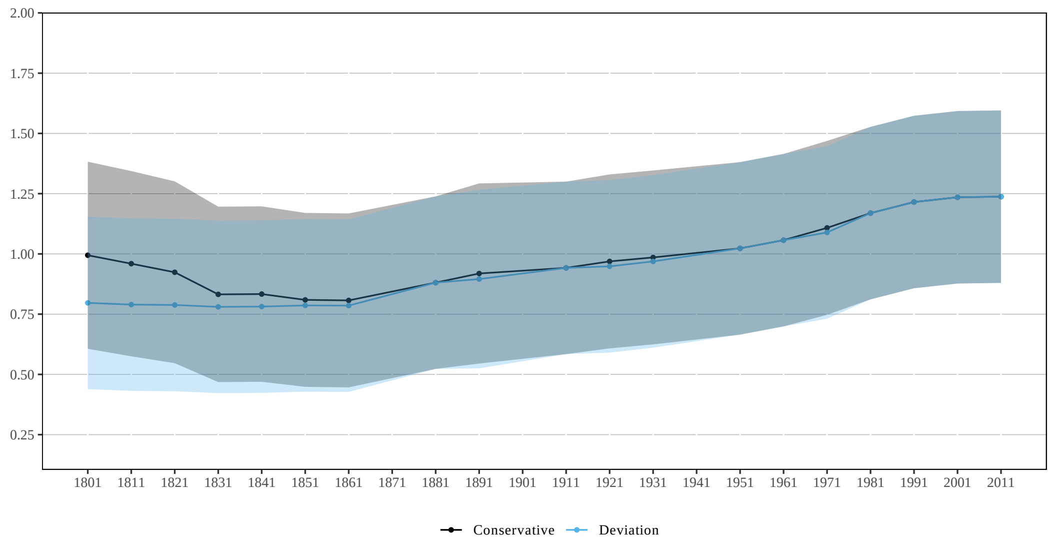
The minimum size of included cities is typically greatest using the *Fraction* cutoff: only for LGDs, the smallest units, before 1961 and for UAs (before 1831) is this not the case – and in those instances, it is the *Level* cutoff that has the largest minimum city size. Given the less arbitrary nature of the *Conservative* and *Deviation* cutoffs, this suggests that a rule-of-thumb *Fraction* cutoff is likely to miss important parts of the city distribution – as seen, for example, in the right-hand panel of Figure 3. Secondly, sample sizes (and related minimum city sizes) are volatile, even for statistically more robust cutoffs, for spatial units based on administrative boundaries. This is most obvious in Figure 1, where the minimum city size, under the *Fraction* or *Deviation* cutoffs, increases substantially between 1971 and 1991 – in a way completely inconsistent with Britain's underlying population dynamics.

Ultimately, of the twelve possible unit-method pairs, it is the Primary Urban Area spatial unit, combined with either the *Conservative* or *Deviation* Cutoff, that reflects the two *a priori* desired attributes: a largely stable number of cities to be included in the analysis, and a minimum city size that grows gradually over time, in line with Britain's urban population. Indeed, for PUAs in Census years after 1981, the *Conservative* and *Deviation* cutoffs give the same sample of sixty urban areas, representing approximately 60% of Britain's overall population in 2011. For simplicity, given the nature of its construction means that it is at least as inclusive as the *Deviation* cutoff, our preferred cutoff is the *Conservative* one.

4 Analysis & Results

In this section, we outline the results of the three elements of our analysis. We start by investigating the size of the Pareto exponent, in other words testing whether Zipf's Law holds, and examine the extent to which the conclusion varies by city definition and

Figure 4: Pareto exponent, Deviation and Conservative cutoff for Primary Urban Areas



Note: This figure shows, using two of the four cutoff methods (Conservative and Deviation cutoff), the absolute value of the Pareto coefficient, for each Census year for the preferred unit (Primary Urban Area).

sample cut-off. We complement this by presenting Lorenz curves and Gini coefficients for urban population. We turn finally to investigating whether Gibrat's Law holds, i.e. whether there is any link between initial city size and future growth.

4.1 Pareto Exponent

Our first empirical objective is to examine whether Zipf's Law holds, using the best combination of city definition and city-size cutoff, and how that answer changes when other unit-method pairs are used, including those pairs dominant in the existing literature on the presence of Zipf's Law and the slope of the Pareto distribution of city size and rank. As explained in Section 2.2, in addition to the choice of spatial unit and cutoffs, we also examine the impact on the estimated parameter and statistical significance of corrections for bias and the standard error.

Our strategy involves estimating, for all twelve unit-method pairs, the parameters of the city size-rank relationship for each year for which that unit-cutoff pairing is available (up to a maximum of 19 Census years). To start, we convert the size-rank relationship into logs. As per Gabaix and Ibragimov (2011), correcting for bias in the OLS estimator involves using the following adjusted log-log specification: $log(Rank - 0.5) = \beta_0 + \beta_1 log(Size)$. We report the absolute value of β, with a greater value corresponding to a steeper downward slope, i.e. a smaller largest city and thus a less concentrated spread of the population. Similarly, as explained in Gabaix and Ioannides (2004) and Gabaix (2009), the standard error estimated by OLS is underestimated by a factor of 5, because the ranking procedure makes the residuals positively autocorrelated. Thus, we compare the unadjusted standard error with one where the standard error is given by the following: $SE = \sqrt{\frac{2}{n}}$.

In each of 19 Census years, therefore, there are up to twelve unit-cutoff pairings and four sets of results per pairing: the standard (biased) OLS estimator and the unbiased estimator, each with unadjusted and adjusted SEs. Of the 48 permutations for any given year, we present two – the unbiased estimator, with adjusted SEs, using Primary Urban Areas, with either the *Conservative* or the *Deviation* cutoff – as the most reliable and thus our preferred, based on the discussion in Sections 2 and 2.3, and contrast the results from alternative permutations against this. The Pareto exponent for these two over time is presented in Figure 4, with the appropriate 95% confidence intervals. The preferred combination of data and methods presents three important stylized facts:

1. Firstly, Zipf's Law cannot be rejected in any one of the 19 Census years where we can use the preferred unit (Primary Urban Areas), either of the preferred cutoffs, and the appropriate estimator and standard errors.

2. Secondly, for most of the period (unsurprisingly given the similar samples), the two cutoffs produce very similar results.

3. Lastly, while the estimates are sufficiently imprecise to not rule out an exponent of one each Census year, any trend is upward over time, certainly after 1861: i.e if there is a trend in Britain's city distribution, it is towards a less unequal distribution over the two centuries covered. At its lowest, the exponent is close to 0.75 while its most recent value is closer to 1.25. For an urban system with a fifth-largest city of one million (roughly the population of the Glasgow metropolitan area in the 2010s), the fall in the exponent implies a significantly smaller largest city: from 15.6m to 5.2m.

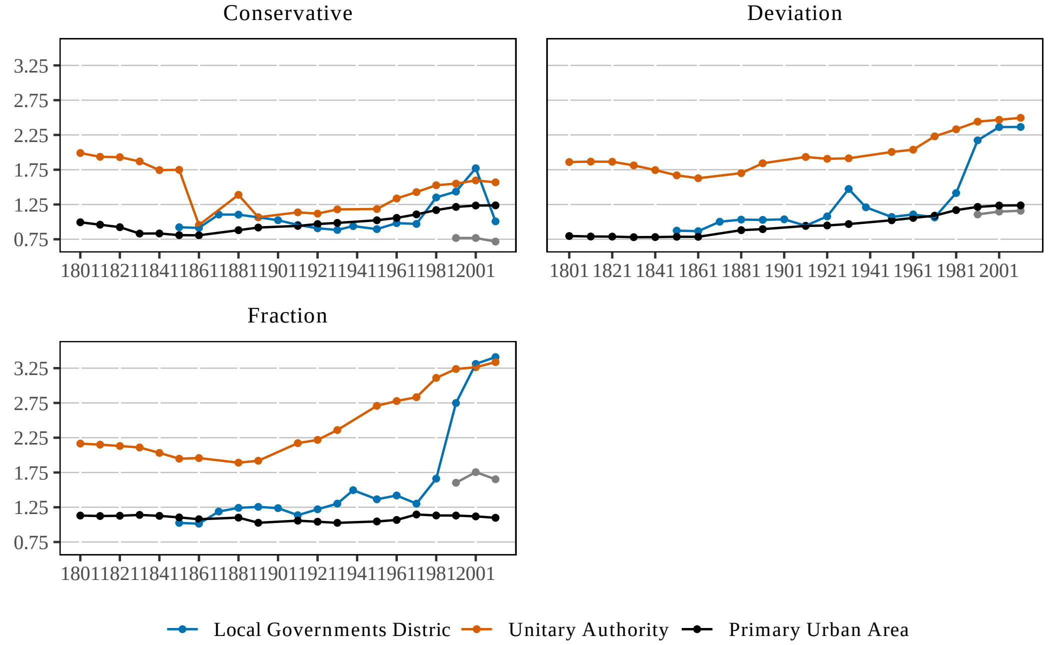
We now explore how these findings are affected firstly by changes in city definition, secondly by changes in sample cut-offs, and finally by the omission of correcting for bias in the coefficient or its standard error. Figures 5 and 6 compare, on panels with standardized axes, the estimated rank-size coefficient across all three definitions of city, for each of the four cutoffs described earlier. In all panels, the coefficients include a correction for bias in the point estimate. For ease of exposition, confidence intervals are not shown but the full set of results is shown graphically in the Appendix, in Figures 14 to 17.

The top two panels of Figure 5 show our preferred cutoffs for city size and the black lines in each are the preferred spatial unit (Primary Urban Areas). As confirmed in 6, there are striking differences between the estimated Pareto exponent for this unit compared to Unitary Authorities and Local Government Districts, the type of administrative units subject to periodic but wholesale revision. Similarly, the use of more ad-hoc cutoffs (such as the top fraction of urban units or a population cutoff) produces very different estimates, in particular for administrative urban units.

Three stylised facts emerge from this comparison. Firstly, across all cutoffs and in almost all periods, these smaller administrative units generate a larger Pareto coefficient. In 2001, using the *Fraction* cutoff, the estimated Pareto coefficient in the LGD and UA datasets is roughly three times that from Primary Urban Areas. While a more extreme example, there are similar results from, for example, the Deviation and Level cutoffs for the most recent Census years.

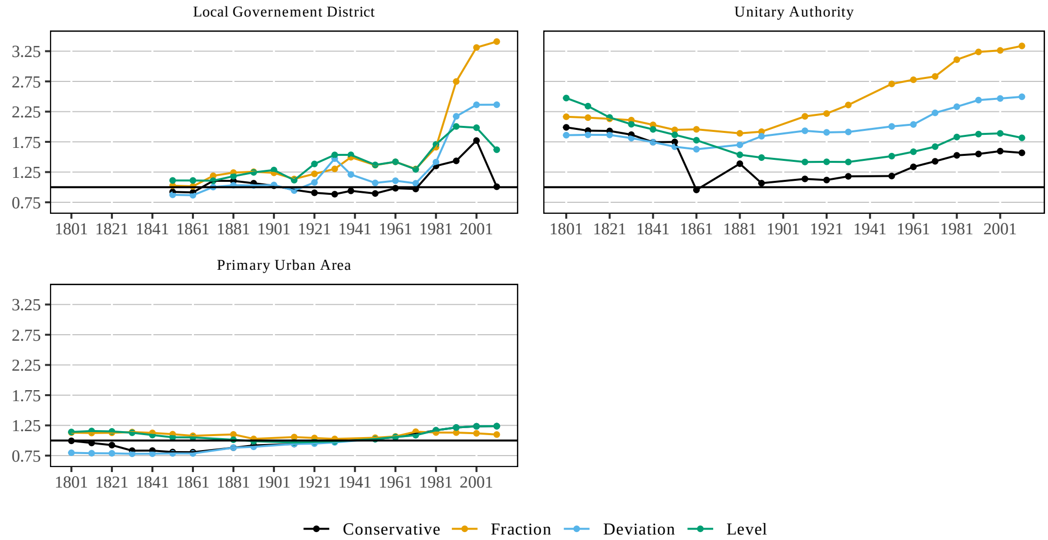
Secondly, using administrative units rather than urban extent, the pattern in the coefficient over time is less consistent within cutoffs. Figure 5 shows, in line with expectations given cities' largely slow-moving populations, a general trend towards a higher coefficient and less concentrated spread of population across cities when Primary Urban Areas are used. Using Unitary Authorities, however, the trend is negative (at least until the mid-20th century) using the *Level* cutoff and more erratic using the

Figure 5: Pareto exponent across spatial units, by cutoff



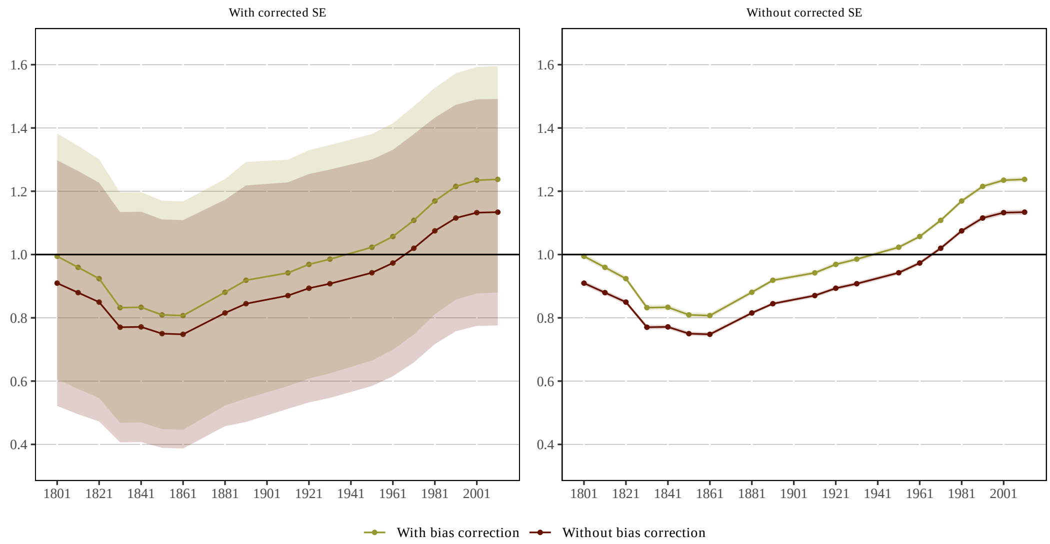
Note: This figure shows, for three definitions of cities, the absolute value of the Pareto coefficient, for each Census year for each cutoff method.

Figure 6: Pareto exponent across cutoff, by spatial units



Note: The four panels in this figure show the comparison of the absolute value of the Pareto coefficient between all cutoff methods, for each Census year and for each unit.

Figure 7: Pareto exponent, with and without correction for bias and standard errors



Note: This figure shows, using the preferred unit-method pair (PUAs and Conservative cutoff), the absolute value of the Pareto coefficient, for each Census year. The left-hand panel shows the results using the appropriate standard errors, while the right-hand panel shows naive standard errors. In both panels, two series are presented: with and without correcting for the bias in estimating the coefficient.

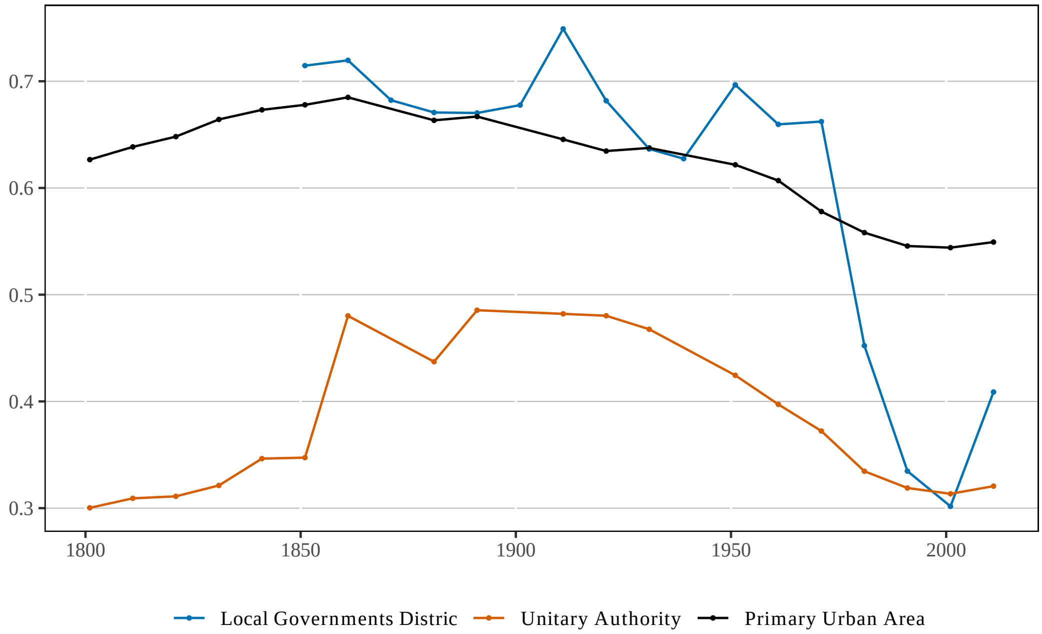
Conservative cutoff. Using LGDs, the trends across the different cutoffs are significantly more erratic, for example, the nearly three-fold increase in the coefficient 1971-2001, using the *Fraction* cutoff.

A related third stylised fact relates to consistency across cutoffs. As mentioned in the paragraph above, where administrative units such as LGDs and UAs are used, the choice of cutoff leads to very different conclusions about the extent of concentration in urban population and its change over time. Focusing just on Local Government Districts in 2011, the *Conservative* cutoff would imply a Pareto coefficient of almost exactly one, the *Level* cutoff almost 1.5, the *Deviation* cutoff 2.25 and the *Fraction* cutoff roughly 3.25.

The final element of our analysis of the Pareto exponent concerns estimation methods. Figure 7 presents for each Census year the estimated coefficient, for the preferred unit-cutoff pairing – Primary Urban Areas and the Conservative Cutoff – with and without bias correction and, in both cases, using adjusted and naive standard errors. The left-hand panel shows the results using the appropriate standard errors, while the right-hand panel shows naive standard errors. In both panels, series with and without correcting for the bias in estimating the exponent are shown.

Two main stylised facts emerge from this. Firstly, the estimated coefficients, when no correction for bias is applied, are larger in absolute value than with the correction applied. This is in line with its construction but it is important to bear in mind, relative to other findings presented in the existing literature: papers without any adjustment for bias in the estimator will understate the extent of urban concentration. Secondly, while naive standard errors imply very precise estimates of the Pareto exponent, the use of appropriate standard errors indicates far greater uncertainty about its value. Again, this follows directly from the construction of the appropriate standard errors but is relevant when considering, for example, the rejection of Zipf's Law (where the exponent equals one).

Figure 8: Gini coefficient across spatial units, *Conservative* cutoff



Note: This figure shows the evolution of the Gini coefficient for our preferred cutoff, i.e. the Conservative cutoff, for each unit.

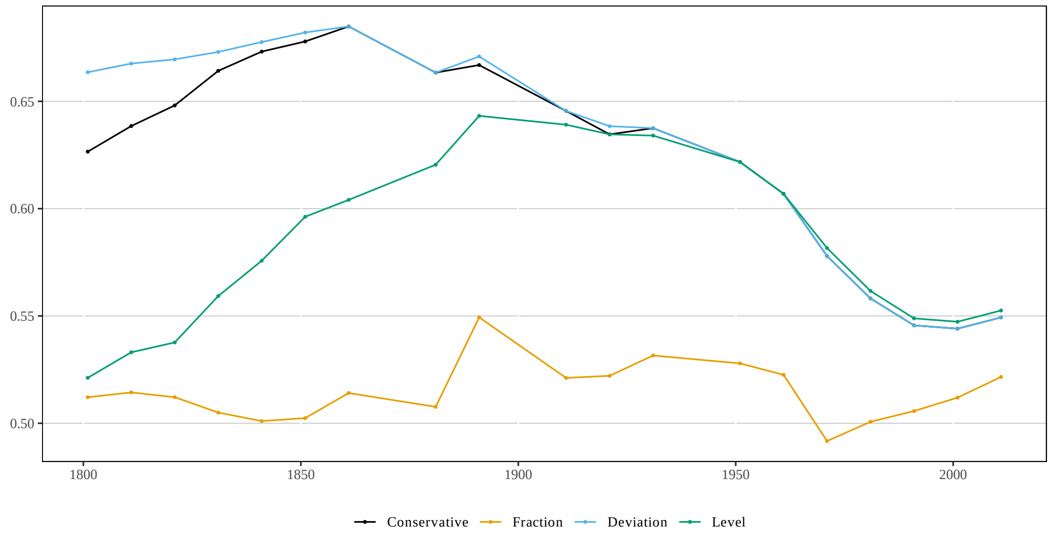
4.2 Gini Coefficient

We turn next to the Gini coefficient of urban population distribution, a measure used by Henderson and Wang (2007) in their analysis of global urban populations. As in its more familiar setting of income distribution, the Gini coefficient ranges between 0 and 1, with 0 representing perfect equality (i.e. all cities have the same population) and 1 representing perfect inequality (i.e. the entire urban population is in one city). As with the Zipf and Gibrat analysis, we compute the Gini coefficient, and associated Lorenz curve, for each of the twelve unit-cutoff pairings, for each Census year for which its available. We present here only the key findings, leaving other analyses for the appendix, in particular Figures 18 and 19 and Tables 7 and 8 for the Gini coefficient and, for Lorenz curves, Figures 20 to 26.

An overview across units, using the Conservative cutoff is given in Figure 8, while an overview across cutoffs, using PUAs is shown in Figure 9. Our analysis presents four stylized facts:

1. Using the preferred combination of unit and cutoff reveals a steady rise and then decline in urban concentration in Britain over time. The Gini coefficient for PUAs rises from 0.63 to 0.69 between 1801 and 1861, a rise of roughly one tenth. The trend then reverses, with that increase undone by 1911, before a pause or significantly reduced fall in concentration between 1911 and 1951. After 1951, the urban population continues to spread, with the Gini coefficient reaching 0.56 in 1991, meaning concentration had fallen by one fifth since 1911. There was little change in concentration between 1991 and 2011. These results point to very different trends in urban concentration in Britain than in those documented by

Figure 9: Gini coefficient by cutoffs, Primary Urban Areas



Note: This figure shows the evolution of the Gini coefficient for our unit, i.e. Primary Urban Area, for each cutoff.

Probst (2017) for Sweden.

2. Figure 8 reveals the importance of choice of urban unit is clear: concentration using Primary Urban Areas is nearly twice that observed in Unitary Authorities. Concentration among Local Government Districts is a similar level and trend to PUAs – but is significantly more volatile and jumps downwards sharply after 1971, reflecting municipality reform.

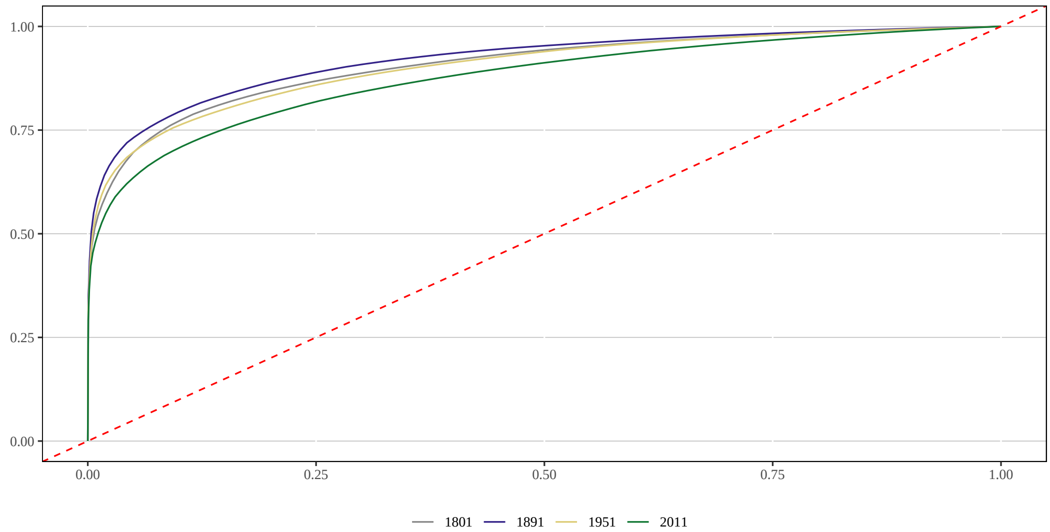
3. Similarly, Figure 9 underscore the effect of choice of cutoff. Using a set fraction of the PUAs would suggest little meaningful change in urban concentration over time. A set population threshold, typically the most common choice of cut-off in the literature, would present an exaggerated increase in concentration during the 19th century, and with different timing (peaking in 1891).

4. The patterns of urban concentration observed through the Gini coefficient are consistent with those seen in the Zipf's Law analysis: a rise in the Pareto exponent (in absolute value) denotes a fall in urban concentration, similar to a decline in the Gini coefficient. The trends in the Gini coefficient and Pareto exponent point to the same evolution of urban concentration in Britain: rising concentration 1801-1861 and falling thereafter, especially 1861-1911 and 1951-1991.

4.3 Gibrat's Law

We turn, finally, to examining whether Gibrat's Law is observed in the data. Gibrat's Law holds that cities follow a growth process that is independent of their size. As per Section 2, Gabaix (1999b) proves mathematically that Gibrat's Law implies Zipf's Law and Córdoba (2008) proves that Zipf's Law implies Gibrat's, but as discussed in Section 2.3, while the validity of this law has been extensively examined, there are no

Figure 10: Lorenz curves for PUAs, by selected years



Note: Lorenz curve for Primary Urban Area with Conservative cutoff.

clear conclusions. The most common approaches in the literature is a non-parametric kernel regression or Nadaraya-Waston estimation. A less well-known method, used by Desmet and Rappaport (2017), is a parametric piece-wise linear spline. We employ both methods here, referring the reader to Desmet and Fafchamps (2006) for more on the detail of these methods. We test Gibrat's Law on three city definitions – LGDs, UAs and PUAs – and for all four cutoffs, but focus our results here on the preferred combination of unit and cutoff: Primary Urban Areas with the Conservative cutoff.[8] Results for other combinations are shown in online Appendix B.

Kernel regression The Kernel regression, or Nadaraya-Watson estimator, gives a continuous nonlinear approximation of growth relative to initial population. Where $L_{i,t}$ refers to the log of population for location i in year t, and with decennial data:

$$(L_{i,t+10} - L_{i,t})/10 = \phi_t L_{i,t} + e_{it}$$

An overview of the results of the Kernel regression, in particular the value of the Nadaraya-Watson estimator, is given in Figure 11, for the preferred combination of city-unit and cutoff. For the period as a whole, the coefficient is statistically significant from zero, at odds with Gibrat's Law. In particular, taking start (1801) and end (2011) populations, cities with a population of 20,000 or less (roughly, below log-value 10) grew faster than cities with a larger population in 1801.

The lower four panels in Figure 11 show that this is driven, largely, by the later 19th century. For 1801-1861, there is no link between initial size and subsequent growth; arguably the same is true for 1911-1951. However, especially for 1861-1911 – and also for 1951-2011 – there appears to be an inverse link between initial city size and subsequent growth: the smallest cities grew faster on average. This echoes Glaeser et al. (2014), who find that, before 1860 and after 1970, less populous counties grew faster in eastern

[8] We do not test Gibrat's Law on Travel To Work Area because of the rationale behind their construction: boundaries are redefined in each of the three census years for which data area available, while guidelines for their construction differ.

and central USA, although the timing is slightly different. The contrast between the 1861-1911 panel and the other three periods, in the scale of the coefficient, suggests a unique set of factors at work then, mostly likely rail infrastructure.

Linear spline The piece-wise linear spline involves mapping (log) population into its vector form, such that the coefficient on each spline segment measures the marginal effect of an increase in population size on growth. If growth is orthogonal, as per Gibrat's Law, the coefficients of each of the spline segments should be close to zero. It is computed according to the following equation, where the 1-by-k vector $L_{i,t}$ includes a constant and a spline of population with $k-1$ segments:

$$(L_{i,t+10} - L_{i,t})/10 = \vec{\beta} \cdot \vec{L_{i,t}} + e_{it}$$

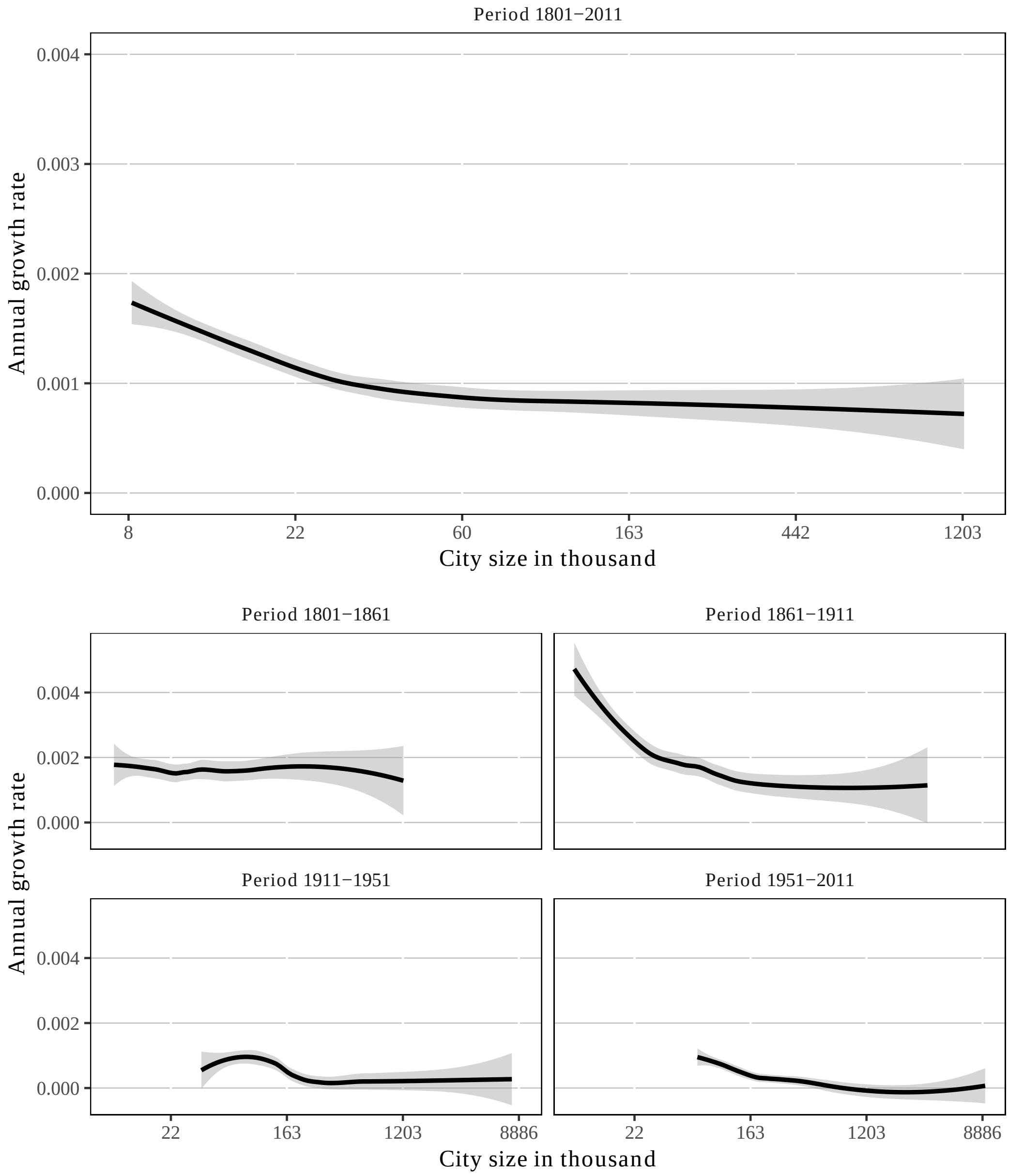
An estimation of the piece-wise linear spline for every unit-cutoff-period combination results in more than 200 regressions, with 819 coefficients. Overall, for 83% (678) of the coefficients, it is not possible to rule out the null hypothesis of no link between initial size and growth, suggesting some support for Gibrat's Law. Rather than report all these coefficients, we instead present a table summarizing the key results for our preferred unit-cutoff pair (see tables 9 to 11).

Tables 9 to 11 summarize the results, for our preferred unit-cutoff pairing. For the nineteenth century, the only significant coefficients are on the smallest size bin for the period 1851-1881. Thereafter, there are a greater number of significant coefficients but the pattern is less clear: smaller cities grow more slowly 1931-1961 and again 1981-2001, while mid-tier cities (with populations of between 0.2m and 1.2m) grow more slowly 1971-2001. Combined, these imply faster growth of the largest cities 1981-2001. These provide more detail in relation to patterns of urban growth, especially in the later 20th century, but are largely consistent with the narrative emerging from the analysis presented earlier in this section.

Taking stock of our analysis of the relationship between initial city size and subsequent growth, we highlight two aspects. Firstly, as with tests of Zipf's Law, conclusions regarding Gibrat's Law depend pivotally on the choice of urban unit. In particular, use of Unitary Authorities, rather than Primary Urban Areas, would strongly imply a negative relationship between initial city size and subsequent growth (see Figure 27 – for the whole period 1801-2011 and for three of the four major sub-periods (1861-1911, 1911-1951 and 1951-2011). Similarly, using LGDs (and the same Conservative cutoff) would lead to a rejection of Gibrat's Law for the period 1961-1981, and in the same direction: smaller LGDs grew faster. Similarly, we contrast our findings with those of Klein and Leunig (2015), who reject Gibrat's Law in every setting, looking at England and Wales during the 19th century; this rejection is driven by parishes with very small populations (under 2,000), well below the threshold for cities suggested by the data.

Secondly, we highlight the consistency in the pattern of results between this analysis and those presented earlier. As with the Zipf and Gini analysis, our Gibrat analysis points to the mid-19th century acting as a turning point in the evolution of Britain's urban system. Before this, smaller cities grew no faster than larger cities, but afterwards – especially 1861-1911 – there is strong evidence of compression in the urban structure.

Figure 11: Kernel regression: Primary Urban Area with Conservative cutoff

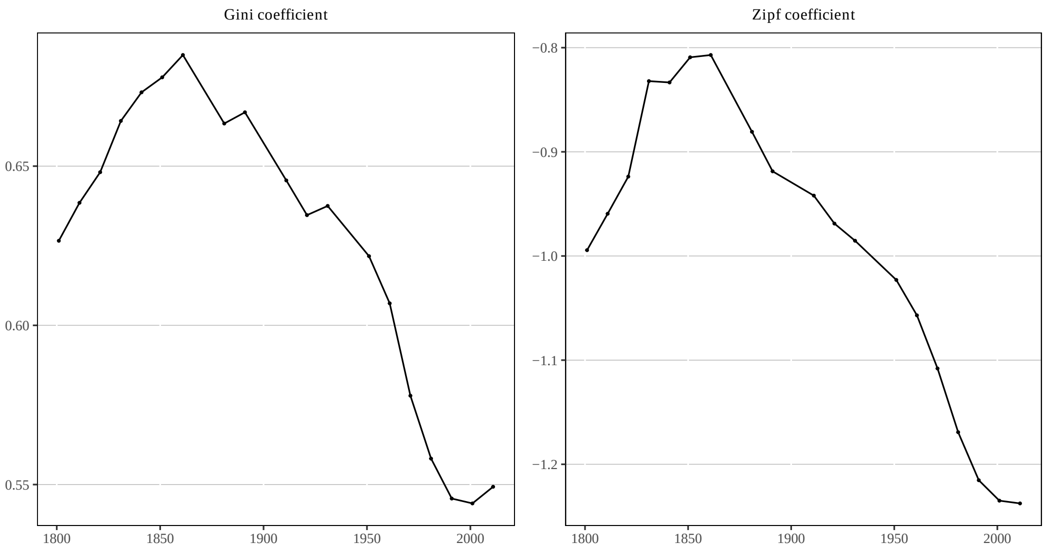


Note: Kernel regression: growth rate for the whole period 1801 to 2011 plotted against initial city size in 1801, and intermediate periods 1801-1861, 1861-1911, 1911-1951 and 1951-2011 plotted against initial size.

5 Conclusion

In this paper, we have examined concentration in Britain's urban system over more than two centuries. In particular, we analysed the relationship between the city size-rank relationship (Zipf's Law) and the link between a city size and its subsequent growth (Gibrat's Law), as well the Gini coefficient as a summary measure of urban concentration. Our work builds on a theoretical literature that has established that naive OLS estimation will produce both a biased estimator and artificially precise standard errors. We also place our work within an extensive empirical literature that examines the city size-rank and size-growth relationships. Using detailed Census data for Britain from 1801, we show how the estimated relationship between a city's rank and size varies dramatically across city definitions and sample cutoffs and that rejection of Zipf's Law depends not only on this choice but also on methods employed. We outline 64 possible combinations but choose Primary Urban Areas, and a Conservative sample cutoff, as our preferred sample, as well as using unbiased estimators and adjusted standard errors. The resulting sample is consistent with *a priori* expectations, given the setting: a largely stable sample size, reflecting the lack of new cities emerging during this period, and a steadily rising minimum city size, reflecting growing urban populations.

Figure 12: Gini coefficient and Zipf coefficient for Primary Urban Area with Conservative cutoff



Note: Gini coefficient and Pareto exponent for Primary Urban Area with Conservative cutoff. Pareto exponent is shown in regular, not absolute, values for consistency across the two panels.

Under our preferred results, reflecting the strongest combination of city definition, sample cutoff, estimator and standard errors, we are unable to reject Zipf's Law, i.e. that the Pareto exponent is one, in any of the 19 Census years available. This reflects the limits to precision as much as the point estimates themselves: the true sample of cities in Britain over the last two centuries is approximately 60. One implication is that in small urban systems, even less precision will be available to researchers. The large standard errors relate, given their formula for construction, to underlying sample size. This suggests that, if the focus is precision of the estimate of the exponent, studies of the distribution should focus on larger economic units. Regardless, it is unclear that we would expect Zipf's Law to hold in economies with very small populations, such as Andorra (2020 population: 77,000) or Iceland (364,000), even if we expect it to hold in larger geographical units, such as Germany (83 million) or Europe as a whole (750m). This raises broader questions about the relevant underlying economic structures: for example, with Britain part of the European Union 1973-2020, is that union the relevant unit within which to understand the distribution of British city sizes in those decades? Similarly, London's very large size in the 19th century may stem from strong links to other parts of the British Empire.

This suggests that it may be helpful for researchers to move the debate on from specifically rejecting (or not) Zipf's Law (whether $\beta = 1$ exactly) and instead understanding patterns of urban concentration (whether β is rising or falling over time). We summarize patterns of urban concentration – across both the Pareto exponent and the Gini coefficient – in Figure 12. Both series show the same pattern over time: a trend towards more concentration in bigger cities between 1801 and 1861 and then the opposite trend thereafter, especially during 1861-1911 and 1951-1991. Both Zipf and Gini analyses suggest a pause in this fall in urban concentration between 1991-2011, the timing of which coincides with the concept of the consumer city, i.e. one based on centripetal forces relating to consumption, rather than production or employment (Glaeser et al., 2001).

The overall change in Pareto exponent is substantial and, where these results have wider relevance, it has significant implications where policymakers wish to understand the likely patterns of urban concentration over coming decades. For example, in the British urban system, the fifth largest city in the earliest 21st century (Glasgow) has a population of close to one million. Where the Pareto exponent is 0.8 (its peak in the mid-19th century), this implies the largest city would have a population of over 15 million. Where the exponent is 1.25, similar to the value seen in 2011, the population of the largest city is less than six million. While the drivers of these changes in urban concentration – including transport technology and policies relating to housing and industrial strategy – are beyond the scope of this paper, the implications are substantial as policymakers seek to accommodate population growth and movements in the 21st century.

In addition, our findings have significant implications for researchers looking to understand the patterns and dynamics of city growth. Across the literature, the modal city definition is legal or administrative, rather than functional, and the principal cutoff used is a fixed population cutoff. If this were used in the case of Britain, it would give, as per Figure 6, an exponent of close to 2 in the early, implying a largest city of just 3 million where the fifth largest is one million.

More generally, across all cutoffs and in almost all periods, the use of smaller administrative units typically generates an artificially large Pareto coefficient, something compounded by the use of arbitrary sample cutoffs, such as a fixed population threshold. In some of the existing empirical literature, this effect may be offset, in part, by use of a naive OLS estimator, which suffers from an attenuation bias. In the case of Britain, the effect is close to 0.1 throughout the two centuries of data. Thus, much of the existing literature may accidentally benefit from two countervailing errors: the lack of adjusting for bias and the use of administrative units, rather than functional cities. In addition, due to the nature of revisions to administrative boundaries over time, these smaller administrative units imply an unrealistic degree of change in urban concentration over time. These issues are not present with the most appropriate unit-method pairs and highlight the need for researchers to understand the spatial units they are analyzing.

Similar to Berry and Okulicz-Kozaryn (2012), we conclude that much of conflict in the literature so far is a consequence of choice of units of observation.

In summary, a growing urban population globally means an understanding of city size and growth patterns is of increasing importance. The experience of Britain over two centuries, the first economy in the world to urbanize, is one of initially rising but then falling urban concentration. In addition, the analysis presented here underscores the importance of meticulous use of data and methods in understanding urban dynamics.

References

Anderson, G., & Ge, Y. (2005). The size distribution of chinese cities. *Regional Science and Urban Economics*, *35*(6), 756–776.

Arshad, S., Hu, S., & Ashraf, B. N. (2019). Zipf's law, the coherence of the urban system and city size distribution: Evidence from pakistan. *Physica A: Statistical Mechanics and its Applications*, *513*, 87–103.

Asadi, S. P. (2019). Empirical investigation of city size distribution in iran. *Available at SSRN 3463117*.

Auerbach, J. (1913). Das gesetz der bevoelkerungskonzentration. *Petermanns Geographische Mitteilungen, 59*, 73–76.

Bairoch, P., Batou, J., & Chevre, P. (1988). The population of european cities. data bank and short summary of results: 800-1850.

Bajracharya, P., & Sultana, S. (2020). Rank-size distribution of cities and municipalities in bangladesh. *Sustainability*, *12*(11), 4643.

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

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

Bettencourt, L. M., Lobo, J., & West, G. B. (2008). Why are large cities faster? universal scaling and self-similarity in urban organization and dynamics. *The European Physical Journal B*, *63*(3), 285–293.

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

Bosker, M., Brakman, S., Garretsen, H., & Schramm, M. (2008). A century of shocks: The evolution of the german city size distribution 1925–1999. *Regional Science and Urban Economics*, *38*(4), 330–347.

Centre for Cities. (2021). *Primary Urban Areas: City definition*. https://www.centreforcities.org/city-by-city/puas/ (accessed: 8 June 2021)

Chauvin, J. P., Glaeser, E., Ma, Y., & Tobio, K. (2017). What is different about urbanization in rich and poor countries? cities in brazil, china, india and the united states. *Journal of Urban Economics*, *98*, 17–49.

Cheshire, P. (1999). Trends in sizes and structures of urban areas. *Handbook of regional and urban economics*, *3*, 1339–1373.

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

Córdoba, J. C. (2008). A generalized gibrat's law. *International Economic Review*, *49*.

Davis, D. R., & Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *The American Economic Review*, *92*(5), 1269–1289.

Decker, E. H., Kerkhoff, A. J., & Moses, M. E. (2007). Global patterns of city size distributions and their fundamental drivers. *PLoS One*, *2*(9), e934.

Desmet, K., & Fafchamps, M. (2006). Employment concentration across us counties. *Regional Science and Urban Economics*, *36*(4), 482–509.

Desmet, K., & Rappaport, J. (2017). The settlement of the united states, 1800–2000: The long transition towards gibrat's law. *Journal of Urban Economics*, *98*, 50–68.

Dingel, J. I., Miscio, A., & Davis, D. R. (2019). Cities, lights, and skills in developing economies. *Journal of Urban Economics*, *31*(6), 103–174.

Dittmar, J. (2020). Cities, markets, and growth: The emergence of zipf's law. *Explorations in Economic History*.

Dobkins, L., & Ioannides, Y. (2001). Spatial interactions among us cities: 1900-1990. *Regional science and urban Economics*, *31*(6), 701–731.
Dobkins, L., Ioannides, Y. Et al. (2000). Dynamic evolution of the us city size distribution. *The Economics of Cities*, 217–260.
Duranton, G. (2007). Urban evolutions: The fast, the slow, and the still. *American Economic Review*, *97*(1), 197–221.
Eaton, J., & Eckstein, Z. (1997). Cities and growth: Theory and evidence from france and japan. *Regional science and urban Economics*, *27*(4), 443–474.
Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, *94*(5), 1429–1451.
Fragkias, M., & Seto, K. C. (2009). Evolving rank-size distributions of intra-metropolitan urban clusters in south china. *Computers, Environment and Urban Systems*, *33*(3), 189–199.
Gabaix, X. (1999a). Zipf's law and the growth of cities. *The American Economic Review*, *89*(2), 129–132.
Gabaix, X. (1999b). Zipf's law for cities: An explanation. *The Quarterly journal of economics*, *114*(3), 739–767.
Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, *1*(1), 255–294.
Gabaix, X., & Ibragimov, R. (2011). Rank- 1/2: A simple way to improve the ols estimation of tail exponents. *Journal of Business & Economic Statistics*, *29*(1), 24–39.
Gabaix, X., & Ioannides, Y. M. (2004). The evolution of city size distributions. *Handbook of regional and urban economics*, *4*, 2341–2378.
Gibrat, R. (1931). *Les inégalites économiques*. Paris : Recueil Sirey.
Giesen, K., & Südekum, J. (2011). Zipf's law for cities in the regions and the country. *Journal of Economic Geography*, *11*(4), 667–686.
Giesen, K., Zimmermann, A., & Suedekum, J. (2010). The size distribution across all cities–double pareto lognormal strikes. *Journal of Urban Economics*, *68*(2), 129–137.
Gisbert, F. J. G., & Mas, M. (2010). La distribución empırica del tamaño de las ciudades en españa, 1900-2001.¿ quién verifica la ley de zipf? *Revista de Economıa Aplicada*, *18*(54), 133–159.
Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of economic geography*, *1*(1), 27–50.
Glaeser, E. L., Ponzetto, G. A., & Tobio, K. (2014). Cities, skills and regional change. *Regional Studies*, *48*(1), 7–43.
González-Val, R., & Lanaspa, L. (2016). Patterns in us urban growth, 1790–2000. *Regional Studies*, *50*(2), 289–309.
González-Val, R., Lanaspa, L., & Sanz-Gracia, F. (2014). New evidence on gibrat's law for cities. *Urban Studies*, *51*(1), 93–115.
González-Val, R., & Silvestre, J. (2020). An annual estimate of spatially disaggregated populations: Spain, 1900–2011. *The Annals of Regional Science*, 1–18.
Henderson, J. V., & Wang, H. G. (2007). Urbanization and city growth: The role of institutions. *Regional Science and Urban Economics*, *37*(3), 283–313.
Ioannides, & Overman, H. G. (2003). Zipf's law for cities: An empirical examination. *Regional science and urban economics*, *33*(2), 127–137.
Ioannides, & Zhang, J. (2017). Walled cities in late imperial china. *Journal of Urban Economics*, *97*, 71–88.
Jiang, B., Yin, J., & Liu, Q. (2015). Zipf's law for all the natural cities around the world. *International Journal of Geographical Information Science*, *29*(3), 498–522.

Klein, A., & Leunig, T. (2015). *Gibrat's law and the british industrial revolution* (No. 221/2015).
Krugman, P. (1996). Confronting the urban mystery. *Journal of the Japanese and International Economies*, *10*(4), 399–418.
Lanaspa, L., Pueyo, F., & Sanz, F. (2003). The evolution of spanish urban structure during the twentieth century. *Urban Studies*, *40*(3), 567–580.
Le Gallo, J., & Chasco, C. (2008). Spatial analysis of urban growth in spain, 1900–2001. *Empirical economics*, *34*(1), 59–80.
Levy, M. (2009). Gibrat's law for (all) cities: Comment. *American Economic Review*, *99*(4), 1672–75.
Matlaba, V. J., Holmes, M. J., McCann, P., & Poot, J. (2013). A century of the evolution of the urban system in brazil, In *Review of urban & regional development studies: Journal of the applied regional science conference*. Wiley Online Library.
Michaels, G., Rauch, F., & Redding, S. J. (2012). Urbanization and structural transformation. *The Quarterly Journal of Economics*, *127*(2), 535–586.
Muller, N. Z. (2016). Power laws and air pollution. *Environmental Modeling & Assessment*, *21*(1), 31–52.
Newman, M. (2005). Power laws, pareto distributions and zipf's law. *Contemporary physics*.
Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics*, *57*(1), 86–100.
Probst, J. (2017). *Agglomeration forces: A study of swedish cities and municipalities since 1800*.
Ronsse, S., & Standaert, S. (2017). Combining growth and level data: An estimation of the population of belgian municipalities between 1880 and 1970. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *50*(4), 218–226.
Rosen, K. T., & Resnick, M. (1980). The size distribution of cities: An examination of the pareto law and primacy. *Journal of urban economics*, *8*(2), 165–186.
Rossi-Hansberg, E., & Wright, M. L. (2007). Establishment size dynamics in the aggregate economy. *The American Economic Review*, *97*(5), 1639–1666.
Rozenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, *101*(5), 2205–25.
Schaffar, A., & Dimou, M. (2012). Rank-size city dynamics in china and india, 1981–2004. *Regional Studies*, *46*(6), 707–721.
Sharma, S. (2003). Persistence and stability in city growth. *Journal of Urban Economics*, *53*(2), 300–320.
Small, C., & Elvidge, C. D. (2013). Night on earth: Mapping decadal changes of anthropogenic night light in asia. *International Journal of Applied Earth Observation and Geoinformation*, *22*, 40–52.
Small, C., Elvidge, C. D., Balk, D., & Montgomery, M. (2011). Spatial scaling of stable night lights. *Remote Sensing of Environment*, *115*(2), 269–280.
Soo, K. T. (2005). Zipf's law for cities: A cross-country investigation. *Regional science and urban Economics*, *35*(3), 239–263.
Soo, K. T. (2007). Zipf's law and urban growth in malaysia. *Urban Studies*, *44*(1), 1–14.
Soo, K. T. (2014). Zipf, g ibrat and geography: Evidence from c hina, i ndia and b razil. *Papers in Regional Science*, *93*(1), 159–181.
Southall, H. (2017). *A vision of britain through time*. https://www.visionofbritain.org.uk/ (accessed: 8 June 2021)
Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, *35*, 40–59.

UK Houses of Parliament. (2018). The Bournemouth, Dorset and Poole (Structural Changes) Order 2018. *Westminster: UK*. https://www.legislation.gov.uk/uksi/2018/648/article/2/made

UN-DESA. (2018). *World urbanization prospects: The 2018 revision*.

Valbuena, G. J. P., & Roca, A. M. (2014). Ley de zipf y de gibrat para colombia y sus regiones: 1835-2005. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History*, *32*(2), 247–286.

Zipf, G. K. (1965). Human behavior and the principle of least effort. cambridge, mass., 1949. *MK Petrov's non-linear thinking model as social sciences gnoseology crisis overcoming project*.

Appendices

A Literature

In this section, we outline the findings of the literature review summarized briefly in Section 2.3 above. In total, we review almost fifty published empirical analyses and to aid exposition, we structure our review in two parts – focusing first on shorter-run studies (either cross-sectional or less than a half-century), which typically test Zipf's Law, before turning to longer-run studies, which are more likely to focus on Gibrat's Law. In the former, we pay particular attention to the definition of city used, as this has implications for the pattern of findings, and in the latter, we group papers by region. Ahead of our review, we highlight that papers published prior to Gabaix and Ioannides (2004) will use naive standard errors, while papers published prior to Gabaix and Ibragimov (2011) will not correct for the bias in the OLS estimator, although some will use the Hill estimator.

A.1 Short-run analyses

We define the "short-run" literature to be empirical analyses using less than 50 years of data on urban populations and review a range papers that meet these criteria, distinguishing where possible between three sets of analyses, reflecting the choice of urban unit. The first uses officially-defined municipalities to delineate cities, the second uses metropolitan areas, while the third uses definitions of cities built from satellite imagery. We start with two seminal papers on the topic and a 2005 meta-analysis.

Two seminal cross-countries studies are Rosen and Resnick (1980) and Soo (2005). Both assemble datasets that cover a larger number of countries (44 and 73 respectively), using data from the late 20th century. In both cases, they find that, while estimates vary substantially by country and definition of city, Zipf's Law is far less likely to be rejected when urban agglomerations are used instead of administrative units. Rosen and Resnick (1980) do not employ any bias or error correction; Soo (2005) corrects for bias but not for standard errors, which may affect his results more generally.

The literature to 2002 is reviewed comprehensively in a meta-analysis by Nitsch (2005). They review 515 estimates of the Zipf coefficient from 29 different studies published between 1925 and 2002. Overall, they find that the Zipf coefficient turns out to be significantly larger than 1 (around 1.1), on average, implying that cities are more evenly distributed than Zipf's Law would predict. Estimates closer to 1 are mostly found in studies for the period post-1900 (average is 1.03 for the period after 1901 - 1950) and for metropolitan areas rather than municipalities. In particular, in the US, one of the largest urban systems, Zipf's Law holds for metropolitan areas. These three papers together already highlight the importance of city definition: where metro areas, not municipalities, are used, it is hard to reject Zipf's Law.

Nonetheless, additions to the literature continue to use municipalities as unit of analysis. This includes Giesen et al. (2010) and Giesen and Südekum (2011). In the former, the authors examine a sample of eight countries and find evidence that the double Pareto lognormal distribution provides a better fit than the simple lognormal and Pareto distributions, at odds with Zipf's Law. However, in the latter, the authors use administratively-defined cities within German regions 1975-1997. Applying a cut-off of 100,000 and a corrected OLS estimation, they find that Zipf's law holds *within* *regions*. Two other papers using municipalities reject Zipf's Law. One is Schaffar and Dimou (2012), who study the evolution of Chinese and Indian cities above 100,000 during the period 1981-2004. While they find the city size-rank distribution is Pareto, and cannot reject Gibrat's Law, they conclude that Zipf's Law is systematically violated for both countries, with the exponent greater than one (and varying over time). Using Iranian data 2006-2016 and cutoffs of between 20,000 and 100,000, Asadi (2019) finds a coefficient higher than predicted by Zipf's Law in the truncated sample.

Eeckhout (2004) employs a cross-section of over 25,000 "places" in the US, officially designated so either by state law or by the federal Census in 2000. He uses an OLS estimator on different cutoff levels, from 42 to 155,000, and shows that the Pareto exponent is very sensitive to the choice of the cutoff level. He then shows that a log-normal distribution, rather than a power law, best fits. In a responding paper, Levy (2009) establishes that, for the largest 150 cities – home to almost one quarter of the US population – the relationship is unequivocally described by a power law. This is relevant where the objects of interest are the largest urban units. We take the key findings from both papers – the importance of cut-offs and the power-law distribution of larger urban units – to our analysis.

Five papers in our review use cities as defined by functional metropolitan areas, rather than administrative municipalities. Using 1970s-1990s data for the USA and France, Duranton (2007) links city size distribution with industrial turnover and finds that second-nature industries prevent small cities from disappearing. His model suggests that the steady-state Zipf's curve is concave, with coefficient below 1 in lower tail and above 1 in upper tail. Bettencourt et al. (2008) analyze labor markets in China, US and Europe and find that the processes relating urbanization to economic development and knowledge creation are common to all big cities belonging to the same urban system. They observe that wealth creation and innovation are associated with an exponent greater than one, while infrastructure is associated with an exponent below one.

Chauvin et al. (2017) compare the US, Brazil, India and China 2000-2010. They find that both Gibrat's Law and Zipf's Law hold in Brazil and the USA, but not in China and India, a finding that is suggestive of spatial equilibrium emerging with economic development. Decker et al. (2007) examine global data on cities, using a combination of metropolitan areas and cities defined as night-light clusters, a method discussed further below. They conclude, in line with Eeckhout (2004), that the full distribution of cities is best fit by a log-normal distribution, while the Pareto distribution only emerges in the upper tail. Similarly, Bajracharya and Sultana (2020) combine both official data on metro areas with cities defined by disaggregated spatial data, in their case the street network. Using the case of Bangladesh 1991-2019, and applying the corrected OLS method, they find that Zipf's Law does not hold: instead, it is smaller when all municipalities are considered and concave at the upper end of the distribution.

We turn, lastly, to the newer literature using cities defined by satellite imagery. An early contribution is by Fragkias and Seto (2009), who use contiguous urban built-up areas in three parts of South China's Pearl River Delta, 1988-1999. They find that urban clusters in metropolitan areas do follow a power law distribution but its parameters oscillate overtime. However, their definition of city is likely very small, as their analyses includes over 5,000 built-up areas across three regions with a population of 21 million.

Rozenfeld et al. (2011) define cities as "maximally connected cluster of populated sites defined at high resolution". They use this new measure on US and British populations, for 2001 and 1981 respectively, and find that the size-rank relationship is well described by Zipf's Law. Small et al. (2011) employ a similar approach at a global scale;

specifically, they use spatially contiguous patches of stable night light over a range of brightnesses corresponding to different intensities of anthropogenic development. Using both OLS and MLE and different brightness cutoff levels, they find Power law exponents in the range 0.95 to 1.11, with the estimated slope varying by brightness cut-off. Overall, they also conclude that Zipf's Law holds for a wide range of developed land areas at both continental and global scales.

In related research, Small and Elvidge (2013) examine night-light patterns in China and India 1992-2005. They find that the size distributions of "lighted cities" are consistent with power laws with exponents near -1. The larger lighted segments are closer to spatial networks of contiguous development than individual cities. This finding is consistent with two subsequent analyses using similar data. Using "natural cities" globally extracted from satellite imagery for 1992, 2001 and 2010, Jiang et al. (2015) test for the presence of Zipf's Law. Using an OLS estimator, they are largely unable to reject Zipf's Law, especially at continental and global levels. Exceptions include Africa, in certain periods, and at country level, where Zipf's Law is violated in certain countries and periods. Dingel et al. (2019) construct lights-based metropolitan areas for US, Brazil, China and India and, similarly, are unable to reject Zipf's Law – but would if they had used administratively defined cities. In the case of US and Brazil, their distribution mirrors the distribution of commuting-based definitions of cities.

A.2 Long-run analyses

The second part of our literature review cover empirical analyses of city growth and size distributions in the long-run, which we define to be over a period of at least four decades. We review these below, separating into four broad categories based on their (principal) region of analysis: global studies; the Americas; Asia; and Europe.

Global Henderson and Wang (2007) assemble data on 1,644 cities (usually metropolitan areas, although especially earlier these definitions can vary) in 142 countries over the period 1960-2000. They employ a cutoff of 100,000, which they say reflects national definitions, and eschew the estimation of a Zipf component, preferring instead to analyse the "spatial Gini" coefficients of population inequality at national level. They highlight the importance of institutional variables, with planned economies associated with a smaller Gini coefficient, as are federal and democratic political systems. They also document the importance of "new" cities (i.e. those that grow above the cut-off of 100,000) in driving urban population growth. Related, they find no evidence of concentration into so-called mega-cities in the period under analysis.

Soo (2014) examines the Zipf coefficient in three of the world's most populous countries – Brazil, China and India – between 1950 and 2000. He uses population by sub-national unit (such as region or province) rather than city (however defined), but does employ the GI corrections for estimator and standard error. None of the three countries has more than 30 sub-national units, meaning that standard errors on the Zipf coefficient are large. As a result, he is unsurprisingly unable to reject Zipf's Law in any year for either Brazil or China – in India, Zipf's Law is rejected where at least 20 sub-national units (of 27) are included, with the coefficient being closer to zero.

The Americas Ten papers in our review examine long-run city growth dynamics in the USA, including Rose (2006), who combines both global and USA perspectives.

Specifically, he compares the rank-size relationship between cities (MSAs) within the USA and between the 50 largest countries, over the period 1900-2004 and to 2050 using population projections. Using adjusted standard errors but unadjusted OLS estimators, he finds that Zipf's Law holds both within the USA and across countries - with the coefficient for the 50 largest countries rising from -0.78 to -0.99 between 1900 and 2050.

Five papers focus on US city growth during the 20th century, typically 1900-1990 and using Metropolitan Statistical Areas (MSAs). Dobkins, Ioannides, et al. (2000) reject the null hypothesis of parallel growth of cities and find evidence that the Pareto exponent has been decreasing (in absolute value) over time, implying increased concentration towards the upper end of the distribution; Dobkins and Ioannides (2001) add to this that in cities with neighbours, growth rates are closely interdependent. However, using a non-parametric approach, Ioannides and Overman (2003) find evidence in favor of both Zipf's Law and Gibrat's law: for MSAs, it is not possible to reject the hypotheses that the first two moments of MSA growth, as well as the Pareto exponent, are invariant to city size. Similarly, Black and Henderson (2003) document a stable size distribution and transition process, with bigger cities exhibiting minimal downward mobility. Using a later end date (2010), a higher cutoff (500,000 rather than 50,000), and BEA "Economic Areas", Berry and Okulicz-Kozaryn (2012) find that, when urban regions are properly defined, US urban growth obeys both Gibrat's and Zipf's Law. They conclude that conflict in the literature is a consequence of choice of units of observation.

Four papers analyze city growth in the US over longer horizons, in most instances using Census data that start in 1790. Batty (2006) examines the population of the 100 largest US cities since 1790, together with data for the UK (1901-2001) and globally, using the Chandler dataset from 430BC. By documenting evidence of deviations from growth by proportionate effect, he concludes that rank–size scaling is far from universal, with "micro-level" dynamics of cities rising and falling over time an important aspect to consider. This is consistent with Glaeser et al. (2014), who explore the dynamics of county growth in eastern and central USA over the period 1860-2000, with some data extending back to 1790. While they find evidence in favour of Gibrat's Law for the sample as a whole, this does not hold for long sub-periods. Before 1860 and after 1970, less populous counties grew more quickly, while between, population growth was regularly faster in more populated areas. While one interpretation is that Gibrat's law is universal only over sufficiently long time periods, another is that Gibrat's law is an artefact of the accidental balancing of centripetal forces, which dominated during the industrial era, and centrifugal forces before and after.

González-Val and Lanaspa (2016) analyze the populations of 190 incorporated places in the USA, 1790-2000, their sample reflecting a population cut-off of 100,000 in the year 2000. They find mixed evidence regarding long-run city growth. On the one hand, the unit root hypothesis underpinning random growth cannot be rejected in most specifications: growth does not depend on initial size. However, there is strong evidence in favour of conditional convergence in growth rates within "clubs", suggestive of "local" mean-reversion within size bins. Lastly, Michaels et al. (2012) investigate the nature of population growth and extend their analysis of the USA, from 1800 to 2000, to include rural areas; they compare this with Brazil 1970-2000. They establish six stylized facts about the dynamics of population and employment, when rural areas are added to the picture, one of which is that Gibrat's Law fails when rural areas are included: agricultural employment growth appears to be decreasing in the initial population density, while in urban settings, employment growth is uncorrelated to initial conditions.

Matlaba et al. (2013) also study Brazil, using a dataset of 185 functionally defined urban areas 1907-2008 and the GI method of estimation. Their principal finding is that the power parameter of the size distribution of the 100 largest urban areas grows over time, approaching unity: Zipf's and Gibrat's Laws became steadily more appropriate descriptions of Brazil's city size distribution during the 20th century. Valbuena and Roca (2014) examine Columbian municipalities that are home to 50% of the country's population, 1835-2005. Using the adjusted rank–size relationship and non-parametric techniques, they are unable to reject both Zipf's law and, from the mid-20th century, Gibrat's law. Their results are consistent with changes in the drivers of Columbia's population growth at both national and regional levels from the 1950s.

Asia In addition to Soo (2014) mentioned above, other researchers have examined the city size-rank relationship in China and India over the long run. Anderson and Ge (2005) analyze China during the period 1949-1999, employing maximum likelihood estimation (MLE) on municipalities and prefecture-level cities with a population of greater than 100,000. They find evidence of a stable city size distribution before the reforms of 1980 but of convergence in growth thereafter. They also suggest that the best-fitting distribution in China is log-normal. Using data on the area of population of walled cities during Ming and Qing era China (1368-1911), Ioannides and Zhang (2017) find the size-rank relationship is well described by Zipf's Law.

Two papers examine the case of Japan over the very long run. Eaton and Eckstein (1997) examine both Japan (925-1985) and France (1876-1990), in the case of Japan focusing on urban areas above 250,000. They find that the relative populations of the top 40 urban areas of France and Japan remained constant during periods of industrialization and urbanization and are described quite well by Zipf's Law. This is at odds with Davis and Weinstein (2002), who analyze Japan during the period 600BC-1998AD. They find that long-run city size is very robust even to large temporary shocks, such as the Allied bombing on Japan during WWII, something inconsistent with random growth rates for cities (Gibrat's Law). They suggest instead that the evolution of Japanese cities over the long run is consistent instead with a hybrid theory of locational fundamentals and increasing returns.

Sharma (2003) examines Census-defined cities in India over the period 1901-1991 and finds that urban population is non-stationary. While the population of cities may be parallel in the long-run, reflecting common long-run growth rates, in the short-run deviations occur, typically reflecting exogenous shocks that take less than a decade to dissipate. Arshad et al., 2019 examine the case of Pakistan, across five Census years 1951-1998, using administrative boundaries, including metropolitan and municipal corporations, and OLS methods. They observe that Zipf's law does not hold in any of the five census years at national level but that it is more likely to hold for the city-size distribution at province-level, of which there are four. Soo, 2007 examines the case of Malaysia across five Censuses between 1957 and 2000, using urban areas of at least 10,000 people and the OLS and Hill estimators. For the full sample, Zipf's law is rejected for all periods except 1957, in favor of a more unequal distribution, while in the upper tail, the results better fit Zipf's Law.

Europe European city growth dynamics have also been the subject of a substantial literature, including Eaton and Eckstein (1997) mentioned above, who found evidence in favour of Zipf's Law in France and Japan. Using the Bairoch et al. (1988) dataset of European cities with a population of more than 5,000 over the course of the second millennium, Dittmar (2020) establishes the emergence of a power-law distribution of city size and rank, first in Western Europe (by 1500) and later in Eastern Europe. This is consistent with technological improvements relaxing the land constraint.

Lanaspa et al. (2003) examine the evolution of Spanish urban structure during the twentieth century. They find divergent growth before 1970 and convergent growth 1970-1999, with significant intra-distribution movements. Their OLS analysis includes the top 100-700 cities, legally defined, in Spain. Le Gallo and Chasco, 2008 undertake an analysis of Spanish towns for the period 1900-2001, with cut-offs of 10,000 and 50,000. They find two main phases, one of divergence (1900–1980) and latterly convergence (1980–2001), and also evidence of the influence of the geographical environment on urban population dynamics. Gisbert and Mas (2010) also employ OLS analysis, in their case on Spain's municipal populations 1900-2001. They find that rejection of Zipf's law depends on the concept of cities used.

González-Val et al. (2014) compare Italy, Spain and the USA over the course of the 20th century, using municipality-level data for both Italy and Spain (and incorporated places in the USA). They use the Nadaraya-Watson method and employ a population cutoff of 200. They observe divergent city growth; however, they also find that data are well fitted by a log-normal distribution and that Gibrat's law holds, at least for certain samples. Lastly, in the case of Spain, González-Val and Silvestre (2020) present annual estimates of population for provinces and provincial capital cities in Spain, 1900–2011. Unlike when data from decennial censuses are used, an analysis of their annual series cannot reject Zipf's Law after the 1940s.

Research has also been undertaken on the long-run dynamics of city growth in Belgium, Germany and Sweden, with results largely at odds with Gibrat's Law. Ronsse and Standaert (2017) estimate the population of Belgian municipalities at annual frequency for three sub-periods within the overall period 1880-1970. They reject Gibrat's Law for Belgium in this period, using the Nadaraya-Watson method. Bosker et al., 2008 examines the urban population in Germany, 1925-1999, using administrative city definitions and cut-offs of 50,000 and 100,000. In addition to World War 2 having a major and lasting impact on city size distributions, he finds that city growth is trend-stationary, consistent with increasing returns to scale but at odds with Gibrat's Law of proportional effect. A working paper by Probst (2017) examines municipalities in Sweden 1800-2010, using Census records. They find that Gibrat's Law is rejected in the sample, earlier because of the growth of smaller locations and later because of city agglomeration. The "Zipf" coefficient reaches its peak at 1.15 in 1900 then falls to 0.89 in 2010.

Finally, the paper most closely related to ours in setting is Klein and Leunig (2015). They examine the dynamics of urban growth in England (a large subset of Britain) during the Industrial Revolution period, 1761-1891. They combine data at the parish level to form over 10,000 'recognisable towns' for the Census years 1801-1891 and use data on nearly 600 administrative units known as "hundreds" prior to this. However, they do not appear to employ a cutoff, with the result that their dataset is weighted heavily towards smaller municipalities; in 1895, their mean location has a population of less than 2,500. With this dataset and using the Nadaraya-Watson non-parametric method, they find that Gibrat's Law is violated consistently, although violations of Gibrat's Law are driven by areas with a population of less than 2,000. They also find evidence that large places grew too quickly to be consistent with Gibrat's Law before 1841, especially in locations where the Industrial Revolution took places. The authors do not present any estimates of the Zipf coefficient.

Table 1: Review of Key Papers in the Literature

Author	Method	Cutoff	Country	Time Period	Unit of analysis
Rosen and Resnick, 1980	OLS	Top 50 or 100,000	Cross section of 44 countries	1970	Urban agglomeration and cities
Eaton and Eckstein, 1997	OLS	France: 50,000 Japan: 250,000	France and Japan	France: 1876-1990; Japan: 925 and 1985	Urban areas
Dobkins, Ioannides, et al., 2000	OLS	50,000	US	1900-1990	Metropolitan area
Dobkins and Ioannides, 2001	OLS	50,000	US	1900-1990	Metropolitan area
Davis and Weinstein, 2002	OLS	NA	Japan	-600-1998	Regional density and city size
Ioannides and Overman, 2003	Local Zipf exponent	50,000	US	1900-1990	Metropolitan area
Eeckhout, 2004	OLS	Different cutoffs between 42 and 155,000	US	2000	Legally incorporated ”places” or Census Designated Places
Soo, 2005	OLS; Hill	10,000	Cross section of 73 countries	1971-2001	Urban agglomeration and cities
Batty, 2006	OLS	NA	US and GB	USA: 1790-2000; GB: 1901-2001; World: 430BD-2000AD	US: “cities and other urban places”, GB: “urban places”.
Levy, 2009	OLS	NA	US	2000	Legally incorporated ”places” or Census Designated Places.
Rozenfeld et al., 2011	OLS	NA	US and GB	1981 for GB and 2001 for US	“city” defined as a maximally connected cluster of populated sites defined at high resolution.
Dittmar, 2020	Theil estimator	5,000	Europe	1300-1800	Urban agglomerations
Ioannides and Zhang, 2017	OLS	1000, 800, 400, 200	China	1368-1911	Walled cities area and population
Dingel et al., 2019	OLS	NA	China India and Brazil	2017	Metropolitan areas, reconstructed with satellite images
Giesen et al., 2010	Hill, OLS		DE, US, FR, Brazil, CZE, Hungary, IT, CH	2000	“Places” for US, “cities” for DE, “communes” for FR
Giesen and Südekum, 2011	OLS	100,000	Germany	1975-1997	City proper data within regions
Small and Elvidge, 2013	OLS		Asia	1992-2009	Aggregation from satellite images of lighted areas
Le Gallo and Chasco, 2008	OLS, SUR	10,000 and 50,000	Spain	1900-2001	Towns and municipalities
Berry and Okulicz-Kozaryn, 2012	OLS	500,000	US	1900-2010	Economic Areas (EAs) defined by the Bureau of Economic Analysis of the US Department of Commerce
Fragkias and Seto, 2009	OLS	One isolated urban pixel	China	1988-1999	Contiguous urban built-up areas
Soo, 2007	OLS, Hill	10,000	Malaysia	1957, 1970, 1980, 1991, 2000	Urban areas
Jiang et al., 2015	Hill		World	1992, 2001, 2010	Natural cities extracted from satellite imagery

Table 2: Summary statistics for the four units

	Local Government District			Unitary Authority			Travel to Work Area			Primary Urban Area		
Year	N	Mean	Median	N	Mean	Median	N	Mean	Median	N	Mean	Median
1801				379	26,777	19,894				61	57,273	24,334
1811				379	30,390	22,860				61	68,253	27,972
1821				379	36,038	26,294				61	83,687	32,779
1831				379	41,698	30,373				61	103,386	40,476
1841				379	47,597	34,897				61	124,253	46,980
1851	527	20,910	5,109	379	53,520	38,117				61	148,331	52,238
1861	581	22,679	5,350	380	60,864	42,451				61	176,064	68,048
1871	930	14,864	4,698									
1881	955	18,086	6,017	380	78,109	49,224				61	269,204	111,689
1891	998	20,903	6,546	348	83,301	51,926				57	307,997	131,258
1901	1,110	22,429	6,969									
1911	1,128	32,217	7,936	380	107,451	62,546				61	409,624	177,354
1921	1,143	30,135	8,595	380	112,550	65,388				61	429,480	183,906
1931	1,077	28,424	9,520	380	117,883	69,578				61	459,634	205,284
1941	986	36,927	14,914									
1951	987	47,328	15,549	380	128,564	81,923				61	490,654	220,001
1961	987	48,473	17,462	380	134,810	90,364				61	507,679	244,559
1971	895	53,917	18,698	380	152,092	101,168				61	565,390	289,065
1981	461	139,827	93,292	380	138,840	105,428				61	479,426	268,653
1991	459	114,087	93,146	380	144,473	114,052	243	195,865	87,173	61	487,358	283,121
2001	376	138,409	112,797	380	150,273	119,207	243	108,980	53,298	61	504,148	280,756
2011	347	131,776	97,099	380	161,503	125,499	218	287,873	152,508	61	546,837	276,786

Table 3: Local Government District

Minimum city size					Sample size				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1851	2,418	13,050	635	50,000	1851	462	106	523	34
1861	2,621	14,712	844	50,000	1861	483	117	574	38
1871	3,641	11,446	1,516	50,000	1871	574	187	837	33
1881	3,840	14,941	2,178	50,000	1881	670	191	829	46
1891	4,322	18,026	3,364	50,000	1891	676	200	768	62
1901	3,955	18,929	4,268	50,000	1901	797	223	763	75
1911	3,939	22,551	3,503	50,000	1911	842	226	875	99
1921	3,867	27,118	13,490	50,000	1921	878	228	425	128
1931	3,549	30,754	44,416	50,000	1931	857	216	147	133
1939	4,374	41,621	13,067	50,000	1939	840	198	536	160
1951	4,262	46,270	9,715	50,000	1951	846	198	632	182
1961	6,051	51,336	10,552	50,000	1961	772	198	641	205
1971	6,932	52,175	12,071	50,000	1971	707	179	578	187
1981	10,119	150,086	14,241	50,000	1981	459	93	456	407
1991	9,986	145,884	66,557	50,000	1991	457	93	353	403
2001	24,457	187,908	89,542	50,000	2001	374	75	266	360
2011	7,375	209,122	117,237	50,000	2011	347	70	144	285

Table 4: Unitary Authority

Minimum city size					Sample size				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1801	24,379	36,078	17,904	50,000	1801	155	76	207	45
1811	23,427	41,915	20,795	50,000	1811	184	76	208	55
1821	28,318	48,955	25,040	50,000	1821	184	76	205	72
1831	30,373	57,994	27,298	50,000	1831	190	76	210	97
1841	27,620	64,650	27,620	50,000	1841	226	76	226	118
1851	34,956	70,781	28,217	50,000	1851	208	76	238	133
1861	4,160	81,887	32,235	50,000	1861	366	76	238	157
1881	24,903	100,677	70,725	50,000	1881	309	76	112	188
1891	3,961	105,924	94,701	50,000	1891	344	70	76	185
1911	8,748	152,242	116,538	50,000	1911	373	76	100	245
1921	9,551	166,729	112,113	50,000	1921	375	76	112	255
1931	15,728	179,652	105,014	50,000	1931	375	76	118	278
1951	4,152	209,713	104,299	50,000	1951	379	76	135	321
1961	5,860	210,098	108,208	50,000	1961	379	76	144	338
1971	8,450	213,711	124,330	50,000	1971	379	76	138	355
1981	5,095	195,659	115,288	50,000	1981	379	76	164	359
1991	4,129	202,193	118,089	50,000	1991	379	76	178	364
2001	7,276	210,134	119,126	50,000	2001	379	76	191	366
2011	7,375	231,997	137,835	50,000	2011	379	76	168	370

Table 5: Travel To Work Area

Minimum city size					Sample size				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1991	7,190	258,218	46,975	50,000	1991	243	49	173	170
2001	4,525	158,499	28,122	50,000	2001	238	49	171	126
2011	7,323	379,654	67,363	50,000	2011	216	43	160	171

Table 6: Primary Urban Area

Minimum city size					Sample size				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1801	1,701	58,324	2,055	50,000	1801	61	13	60	14
1811	2,235	65,695	2,503	50,000	1811	61	13	60	17
1821	2,717	82,816	3,281	50,000	1821	61	13	60	19
1831	3,073	108,566	3,784	50,000	1831	61	13	60	25
1841	3,382	124,548	4,629	50,000	1841	61	13	60	28
1851	3,614	137,176	6,413	50,000	1851	61	13	60	34
1861	3,656	162,451	7,793	50,000	1861	61	13	60	38
1881	13,893	257,805	16,667	50,000	1881	61	13	60	46
1891	19,026	306,445	20,503	50,000	1891	57	11	56	47
1911	31,137	345,479	37,246	50,000	1911	61	13	60	58
1921	36,254	350,494	36,789	50,000	1921	61	13	60	58
1931	35,368	358,630	45,396	50,000	1931	61	13	60	59
1951	41,886	381,132	65,139	50,000	1951	61	13	60	60
1961	48,434	429,253	65,776	50,000	1961	61	13	60	60
1971	65,925	470,256	72,023	50,000	1971	61	13	60	60
1981	70,836	444,240	85,154	50,000	1981	61	13	60	60
1991	80,799	457,568	91,909	50,000	1991	61	13	60	60
2001	85,056	452,404	97,568	50,000	2001	61	13	60	60
2011	90,254	479,924	104,640	50,000	2011	61	13	60	60

Table 7: Gini Coefficient

Local Government District					Unitary Authority				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1851	0.715	0.625	0.728	0.603	1851	0.347	0.292	0.363	0.319
1861	0.72	0.632	0.736	0.601	1861	0.48	0.29	0.372	0.338
1871	0.682	0.639	0.712	0.62	1871				
1881	0.671	0.618	0.69	0.606	1881	0.437	0.301	0.341	0.39
1891	0.67	0.6	0.682	0.58	1891	0.485	0.293	0.304	0.4
1901	0.678	0.594	0.673	0.561	1901				
1911	0.749	0.681	0.753	0.667	1911	0.482	0.266	0.298	0.414
1921	0.682	0.568	0.611	0.538	1921	0.48	0.262	0.304	0.413
1931	0.637	0.466	0.422	0.41	1931	0.468	0.248	0.297	0.414
1939	0.627	0.495	0.555	0.489	1939				
1951	0.697	0.598	0.654	0.598	1951	0.424	0.209	0.276	0.385
1961	0.66	0.577	0.631	0.576	1961	0.397	0.203	0.274	0.37
1971	0.662	0.541	0.634	0.543	1971	0.372	0.203	0.256	0.352
1981	0.452	0.45	0.448	0.425	1981	0.335	0.184	0.248	0.319
1991	0.335	0.21	0.278	0.296	1991	0.319	0.176	0.237	0.306
2001	0.302	0.169	0.249	0.287	2001	0.313	0.174	0.236	0.302
2011	0.409	0.161	0.235	0.34	2011	0.321	0.168	0.23	0.311

Table 8: Gini Coefficient

Primary Urban Area					Travel To Work Area				
Year	Conservative	Fraction	Deviation	Level	Year	Conservative	Fraction	Deviation	Level
1801	0.627	0.512	0.664	0.521	1801				
1811	0.638	0.514	0.668	0.533	1811				
1821	0.648	0.512	0.67	0.538	1821				
1831	0.664	0.505	0.673	0.559	1831				
1841	0.673	0.501	0.678	0.576	1841				
1851	0.678	0.502	0.682	0.596	1841				
1861	0.685	0.514	0.685	0.604	1861				
1881	0.663	0.508	0.663	0.62	1881				
1891	0.667	0.549	0.671	0.643	1891				
1911	0.646	0.521	0.646	0.639	1911				
1921	0.635	0.522	0.638	0.635	1921				
1931	0.638	0.532	0.638	0.634	1931				
1951	0.622	0.528	0.622	0.622	1951				
1961	0.607	0.523	0.607	0.607	1961				
1971	0.578	0.492	0.578	0.582	1971				
1981	0.558	0.501	0.558	0.562	1981				
1991	0.546	0.506	0.546	0.549	1991	0.629	0.407	0.544	0.543
2001	0.544	0.512	0.544	0.547	2001	0.591	0.459	0.436	0.462
2011	0.549	0.522	0.549	0.553	2011	0.605	0.446	0.561	0.529

Table 9: Spline regression for Primary Urban Area and Conservative cutoff - 1

	Annual growth rates					
	1801-1811	1811-1821	1821-1831	1831-1841	1841-1851	1851-1861
00to10	0.0004	0.00001	−0.00002	−0.001	−0.001	−0.002*
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
10to11	−0.00002	−0.0004	−0.0001	−0.001	−0.00001	0.001
	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.001)
11to12	−0.0003	0.00002	0.0004	0.001	−0.0001	−0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
12to14	0.001	0.0005	0.0003	0.0004	0.0003	0.0001
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
14to17	−0.297	−0.008	−0.004	−0.003	−0.001	−0.0004
	(0.591)	(0.018)	(0.006)	(0.003)	(0.002)	(0.002)
Constant	−0.002	0.002	0.002	0.007	0.010*	0.016**
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)
Observations	47	49	51	54	55	55
R^2	0.029	0.045	0.038	0.142	0.083	0.080
Adjusted R^2	-0.089	-0.066	-0.069	0.053	-0.011	-0.013

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Spline regression for Primary Urban Area and Conservative cutoff - 2

	Annual growth rates					
	1861-1881	1881-1891	1891-1911	1911-1921	1921-1931	1931-1951
00to10	-0.004^{***}	0.0003				
	(0.001)	(0.004)				
10to11	-0.001	-0.001	-0.001	0.002	0.004	-0.003
	(0.001)	(0.001)	(0.001)	(0.002)	(0.007)	(0.002)
11to12	-0.001	-0.0002	-0.0004	-0.0001	-0.001	-0.001^{**}
	(0.001)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0002)
12to14	0.0001	-0.0003	-0.0002	-0.0003	-0.0003	-0.0002
	(0.001)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0001)
14to17	-0.0004	0.0003	0.0001	0.0001	0.0003	0.0001
	(0.002)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0003)
Constant	0.044^{***}	-0.001	0.003^{***}	-0.001	-0.003	0.004^{**}
	(0.010)	(0.036)	(0.001)	(0.001)	(0.006)	(0.002)
Observations	56	56	55	56	55	56
R^2	0.508	0.139	0.188	0.077	0.157	0.321
Adjusted R^2	0.458	0.053	0.123	0.005	0.090	0.268
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 11: Spline regression for Primary Urban Area and Conservative cutoff - 3

	Annual growth rates					
	1951-1961	1961-1971	1971-1981	1981-1991	1991-2001	2001-2011
00to10						
10to11						
11to12	-0.001^{*}	0.0002	-0.0002	-0.001^{*}	-0.001^{**}	-0.001
	(0.0004)	(0.001)	(0.001)	(0.0005)	(0.0004)	(0.0004)
12to14	-0.0003	-0.001	-0.001^{***}	-0.0005^{***}	-0.0002^{*}	-0.0001
	(0.0002)	(0.0004)	(0.0004)	(0.0002)	(0.0001)	(0.0001)
14to17	-0.00001	-0.0002	0.001	0.0003	0.0002	0.0002
	(0.0004)	(0.001)	(0.001)	(0.0003)	(0.0002)	(0.0002)
Constant	0.001^{***}	0.001	0.0003	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0003)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)
Observations	56	56	56	56	56	56
R^2	0.186	0.073	0.178	0.319	0.207	0.129
Adjusted R^2	0.139	0.020	0.131	0.280	0.162	0.079
Note:						*p<0.1; **p<0.05; ***p<0.01

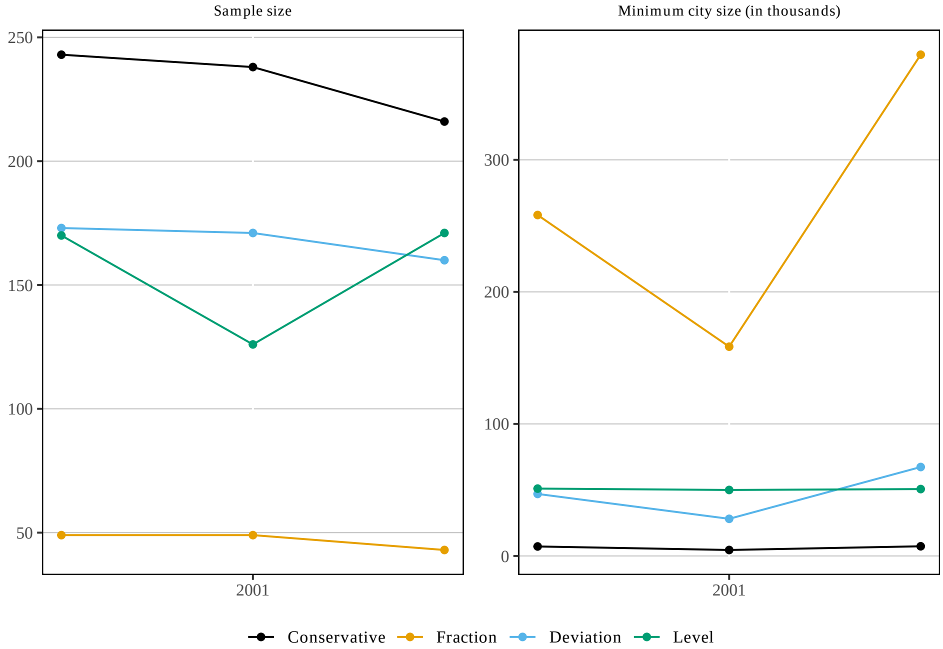
Table 12: Top 5 Primary Urban Area

1801	1861	1911	1951	2011
1,213,042	3,438,830	7,867,873	9,301,193	9,736,823
234,934	928,230	1,902,374	2,247,942	2,419,500
176,878	721,475	1,617,845	1,997,297	1,876,194
141,380	580,330	1,321,586	1,451,969	1,054,473
108,641	490,907	946,181	1,053,225	829,319

Table 13: Bottom 5 Primary Urban Area

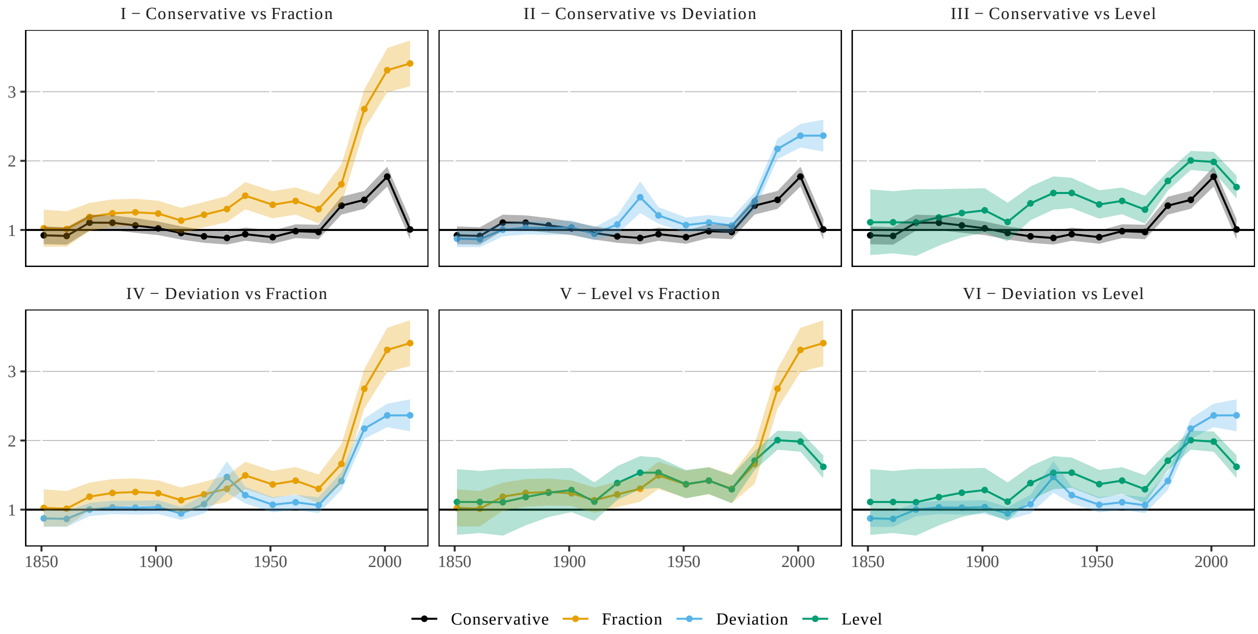
1801	1861	1911	1951	2011
1,701	3,656	31,137	41,886	90,254
2,055	7,793	37,246	65,139	104,640
2,297	10,475	49,645	66,857	121,688
4,162	10,504	53,375	68,021	123,867
4,402	10,566	53,649	70,845	133,384

Figure 13: Travel To Work Area



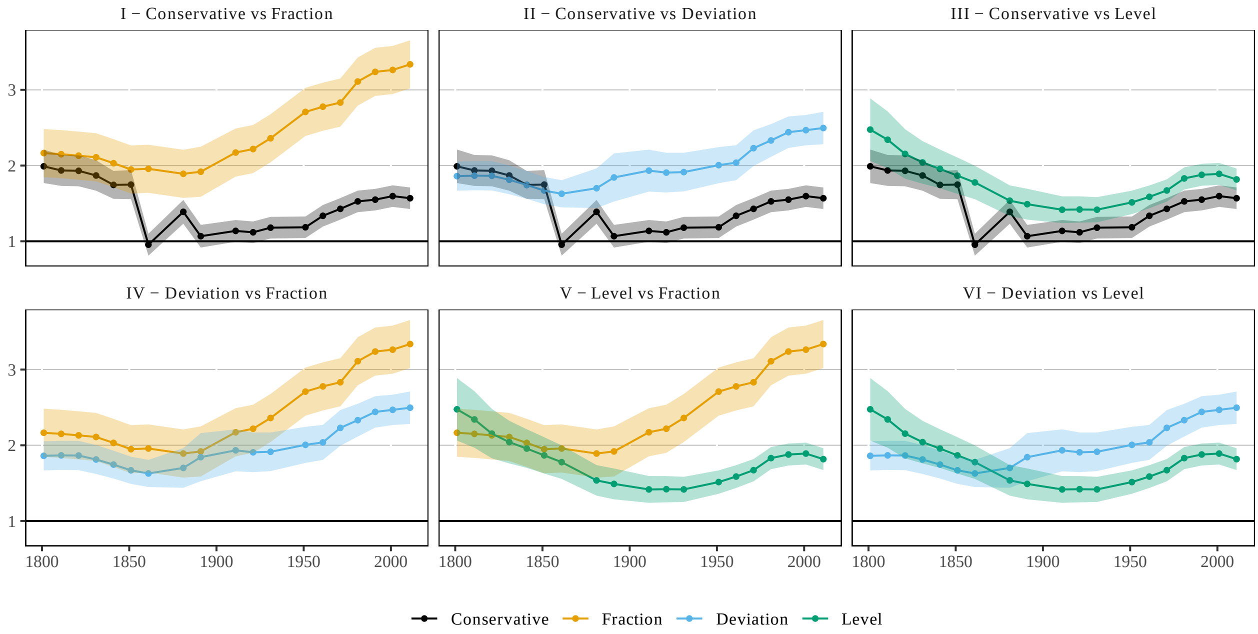
Note: The left panel of this figure shows the number of Travel To Work Area included in the sample for each census year for each cutoff methods. The right panel shows the size of the smallest Travel To Work Area included in the sample for each cutoff method..

Figure 14: Size-rank coefficient: Local Government Districts



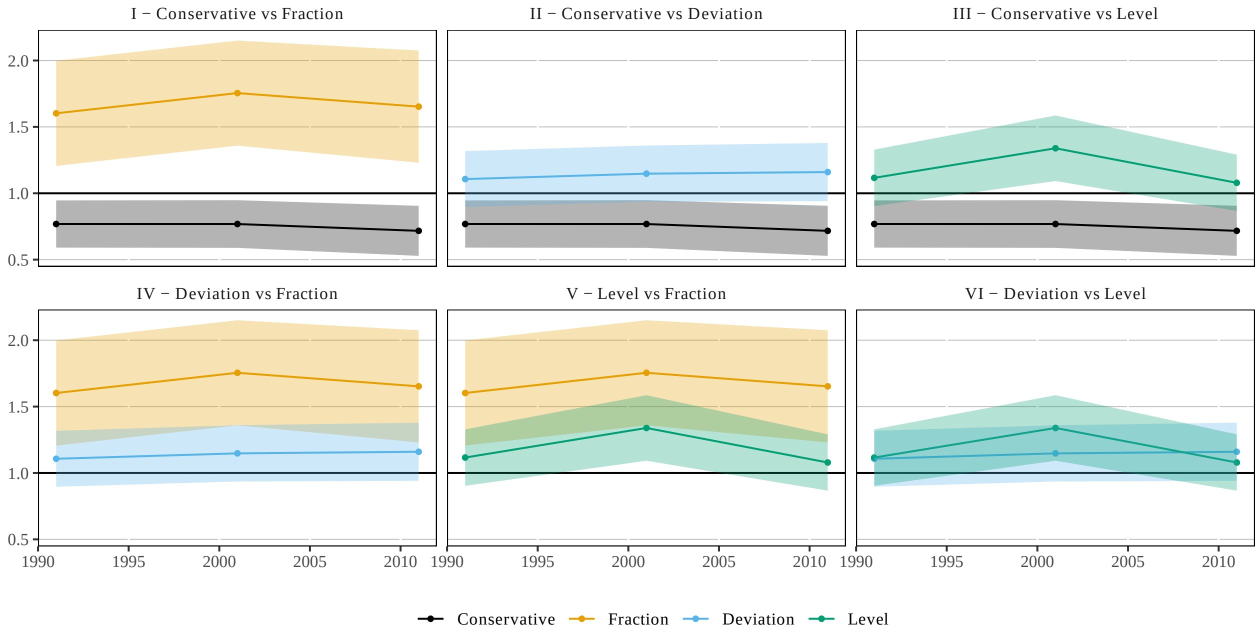
Note: The six panels in this figure show the pairwise comparison between all cutoff methods of the absolute value of the Pareto coefficient of Local Government Districts, for each Census year.

Figure 15: Size-rank coefficient: District/Unitary Authority



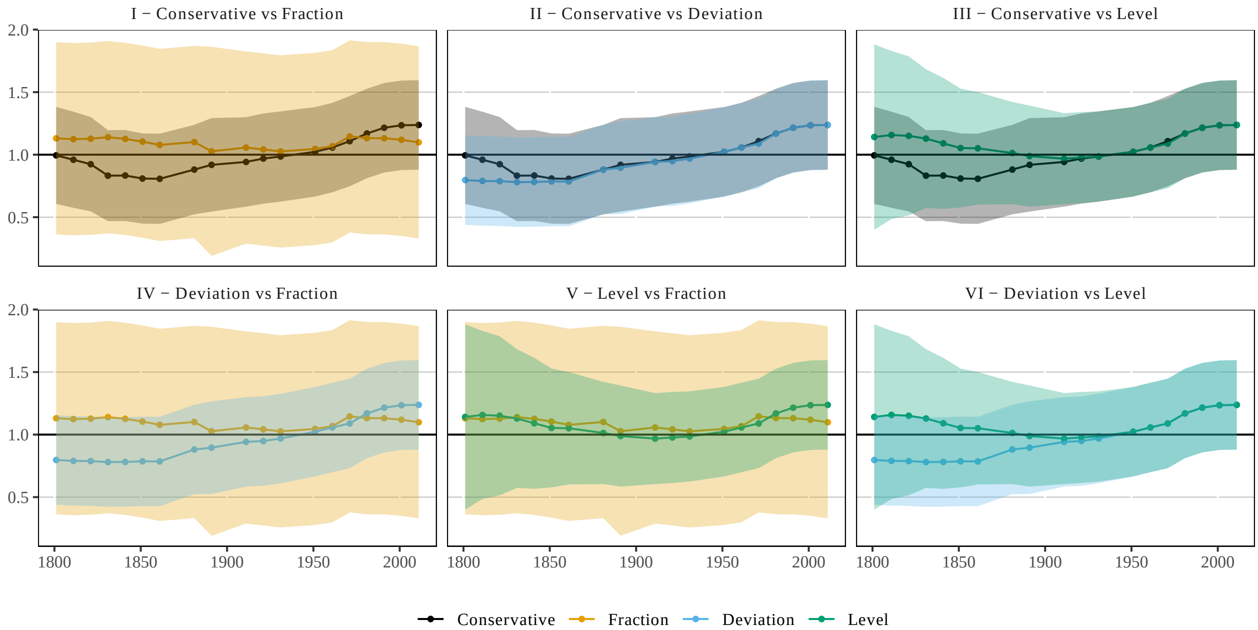
Note: The six panels in this figure show the pairwise comparison between all cutoff methods of the absolute value of the Pareto coefficient of District/Unitary Authority, for each Census year.

Figure 16: Size-rank coefficient: Travel To Work Area



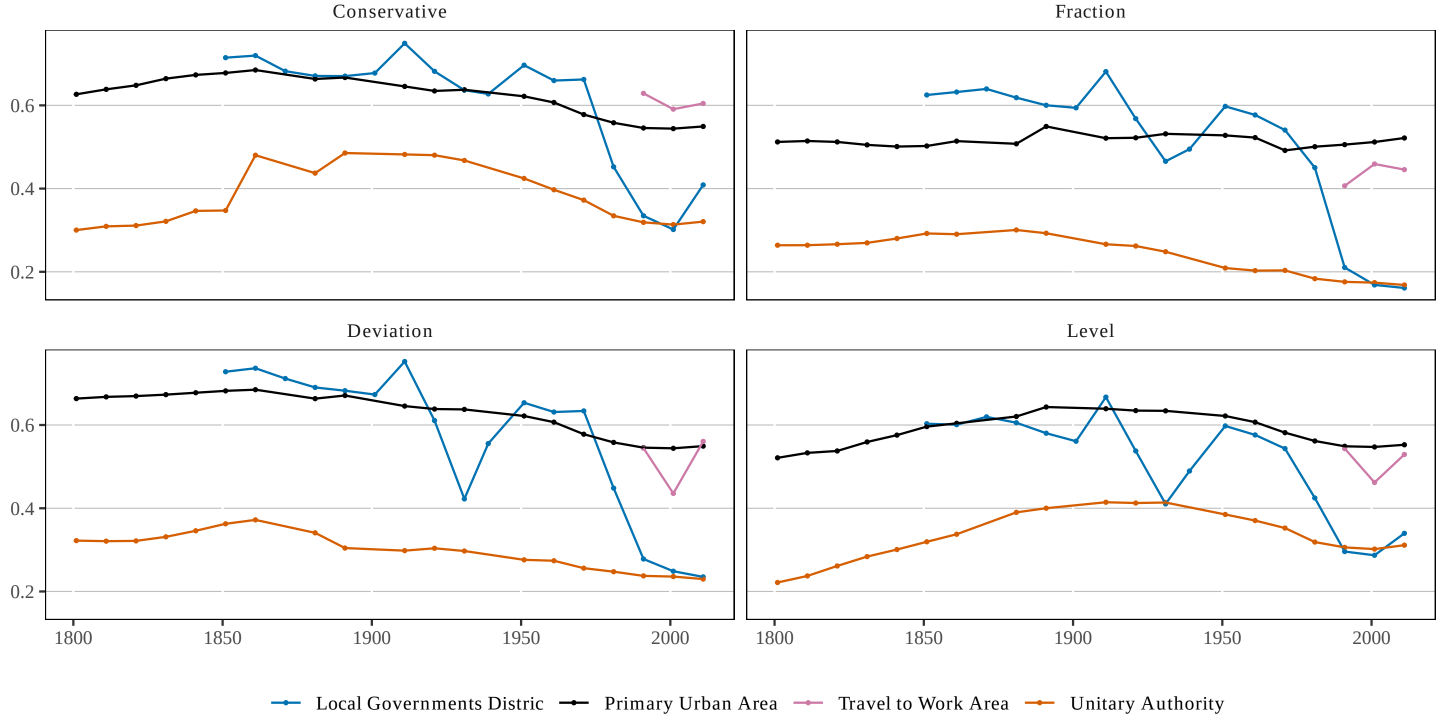
Note: The six panels in this figure show the pairwise comparison between all cutoff methods of the absolute value of the Pareto coefficient of Travel To Work Area, for each Census year.

Figure 17: Size-rank coefficient: Primary Urban Area



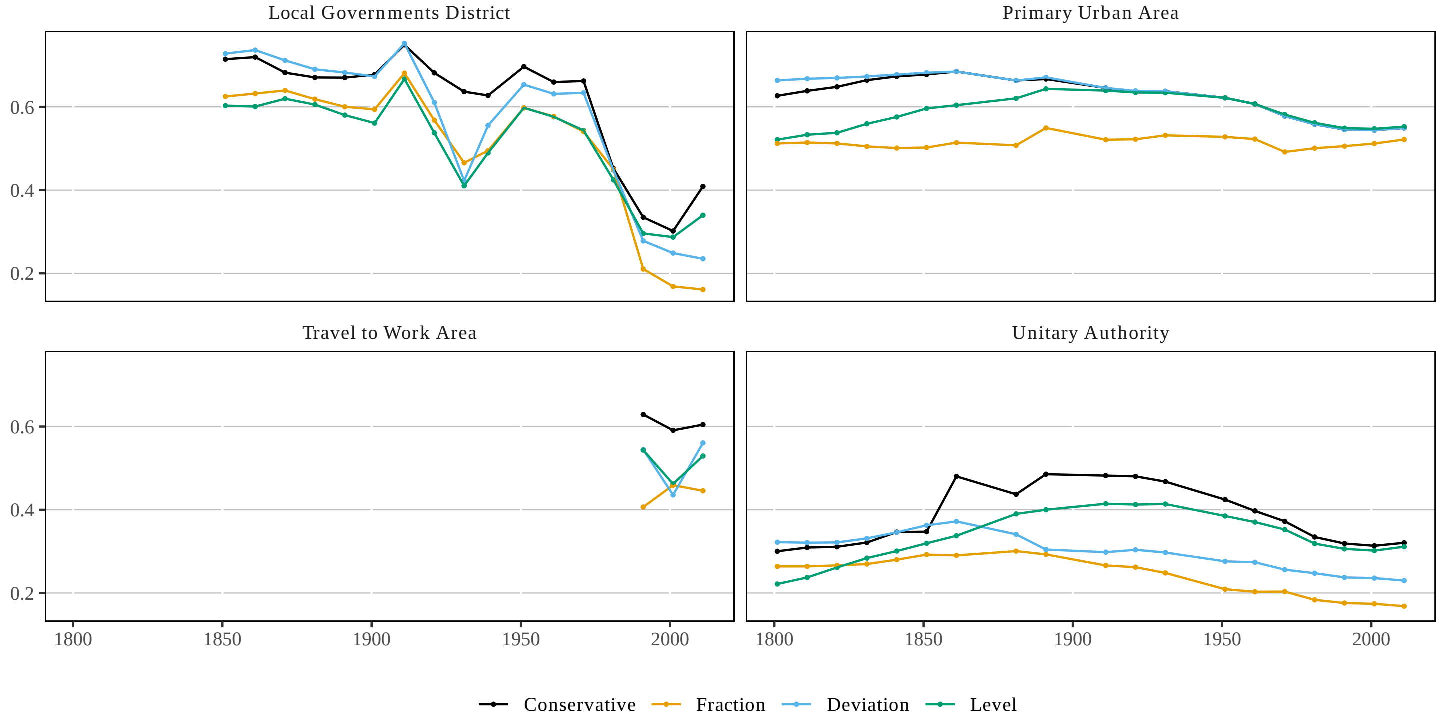
Note: The six panels in this figure show the pairwise comparison between all cutoff methods of the absolute value of the Pareto coefficient of Primary Urban Area, for each Census year.

Figure 18: Gini coefficient for four different units



Note: This figure shows the evolution of the Gini coefficient for all different cutoff in each area.

Figure 19: Gini coefficient for four different cutoffs



Note: This figure shows the evolution of the Gini coefficient for all different units for each different cutoff.

Figure 20: Lorenz curve for Conservative cutoff - Local Government District

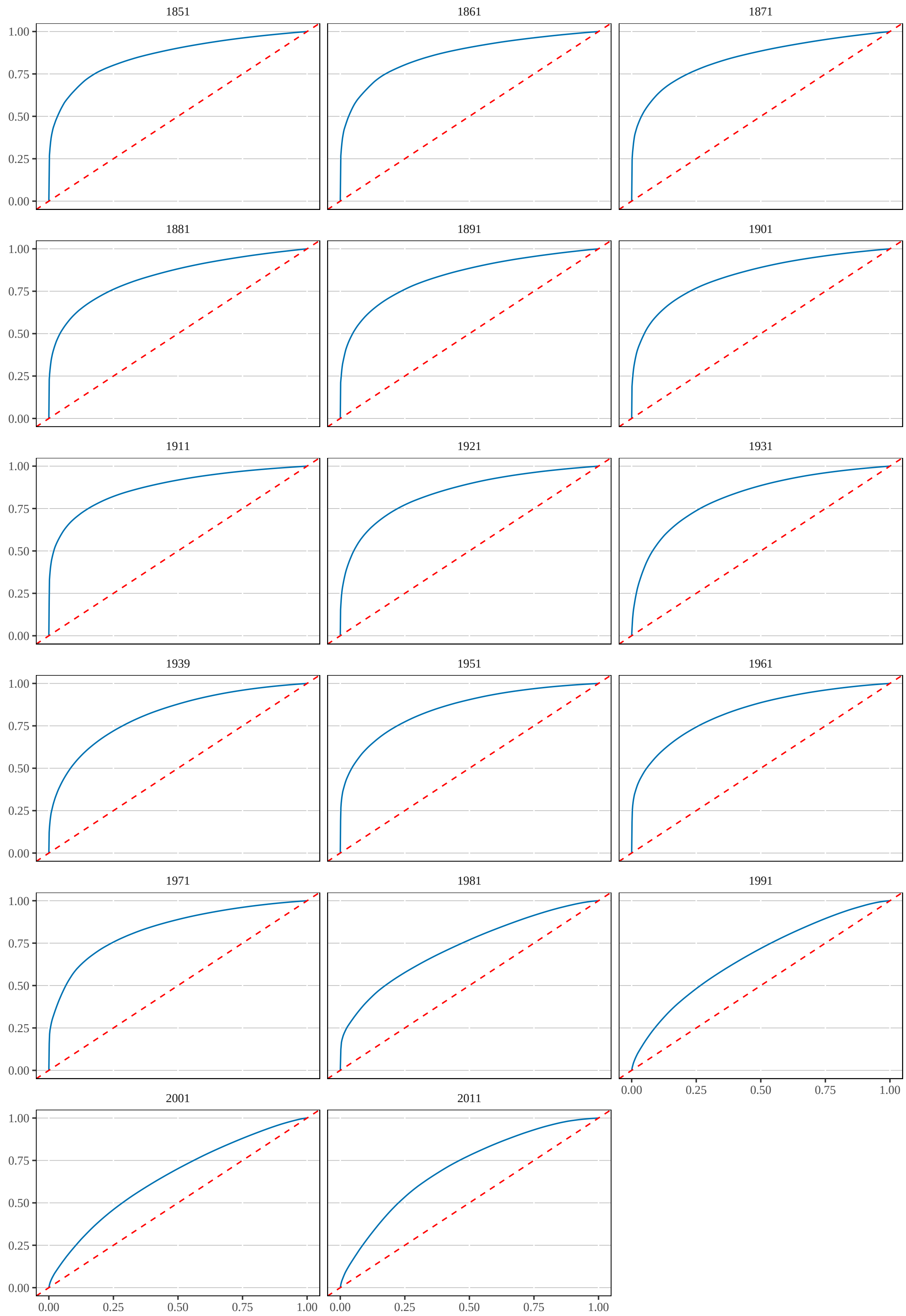


Figure 21: Lorenz curve for Fraction cutoff - Local Government District

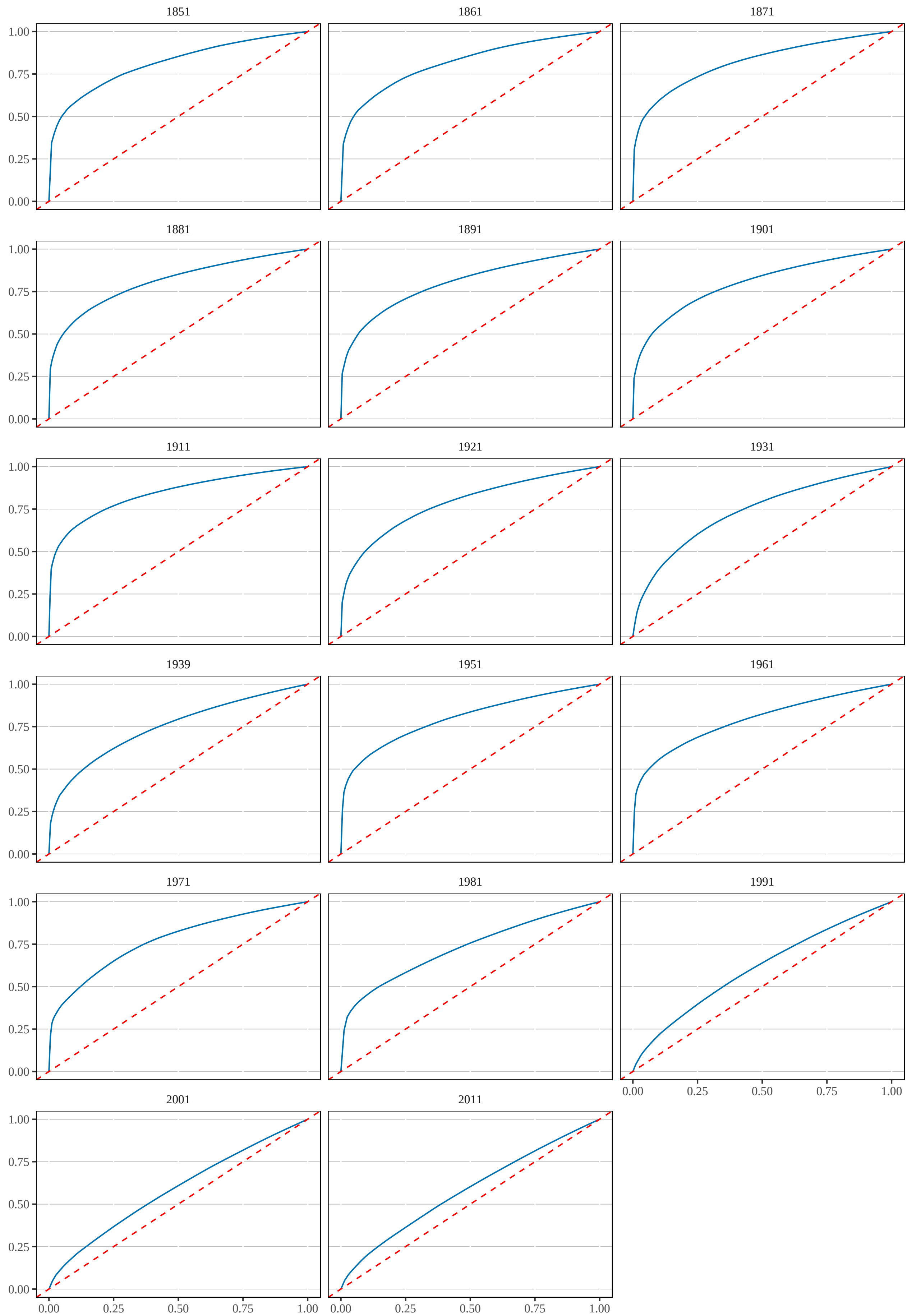


Figure 22: Lorenz curve for Level cutoff - Local Government District

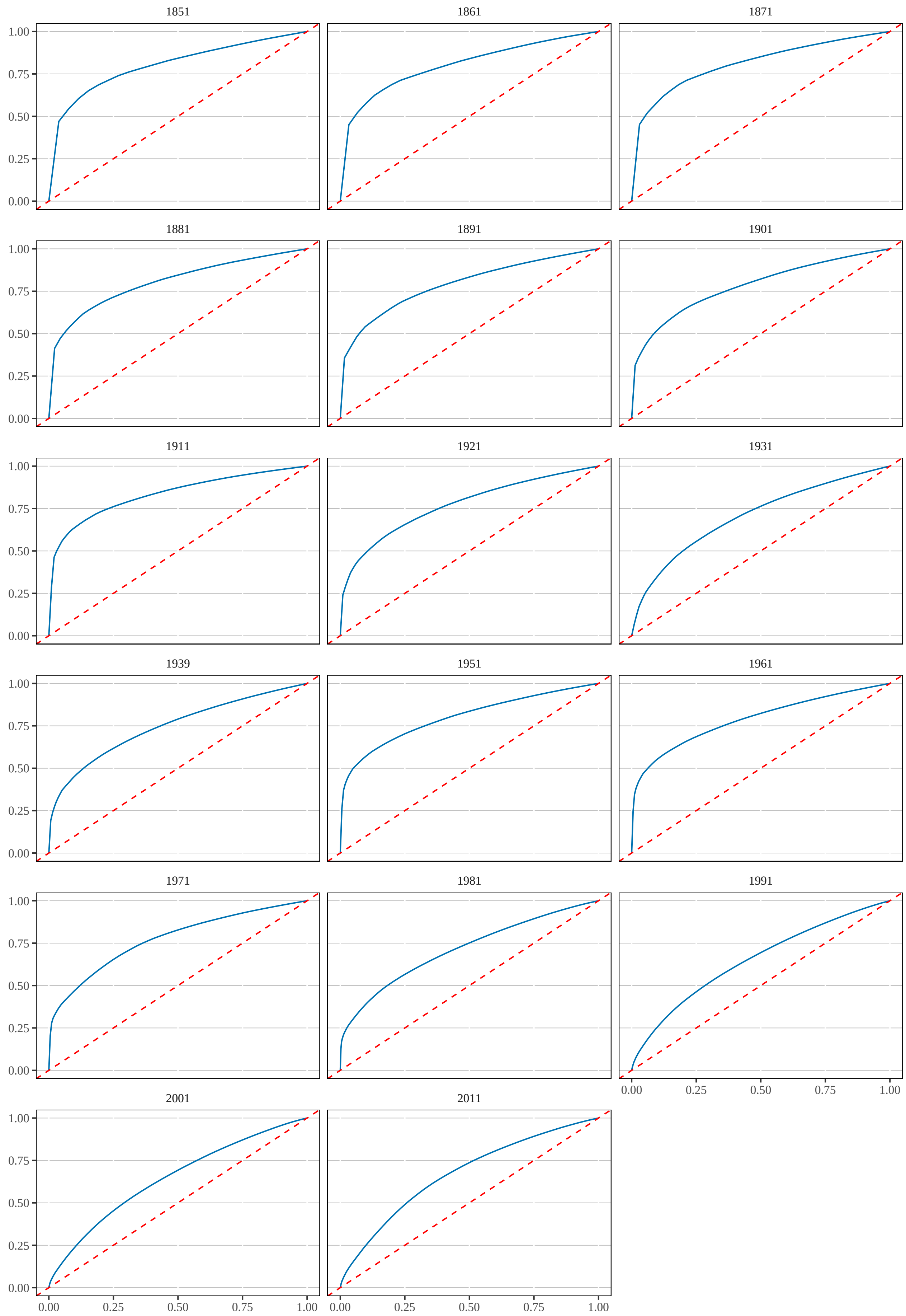


Figure 23: Lorenz curve for Deviation cutoff - Local Government District

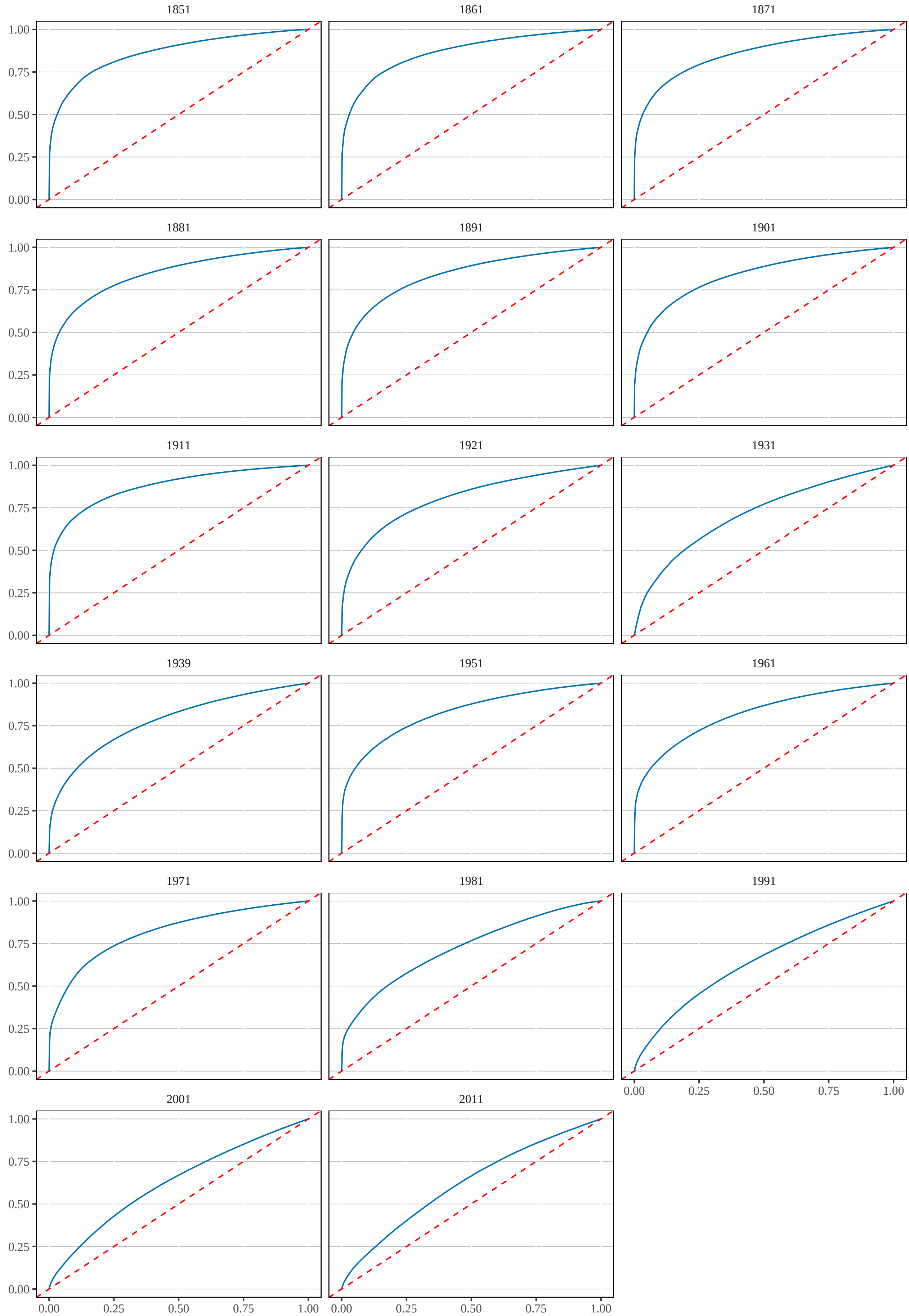


Figure 24: Lorenz curve for Conservative cutoff - Unitary Authority

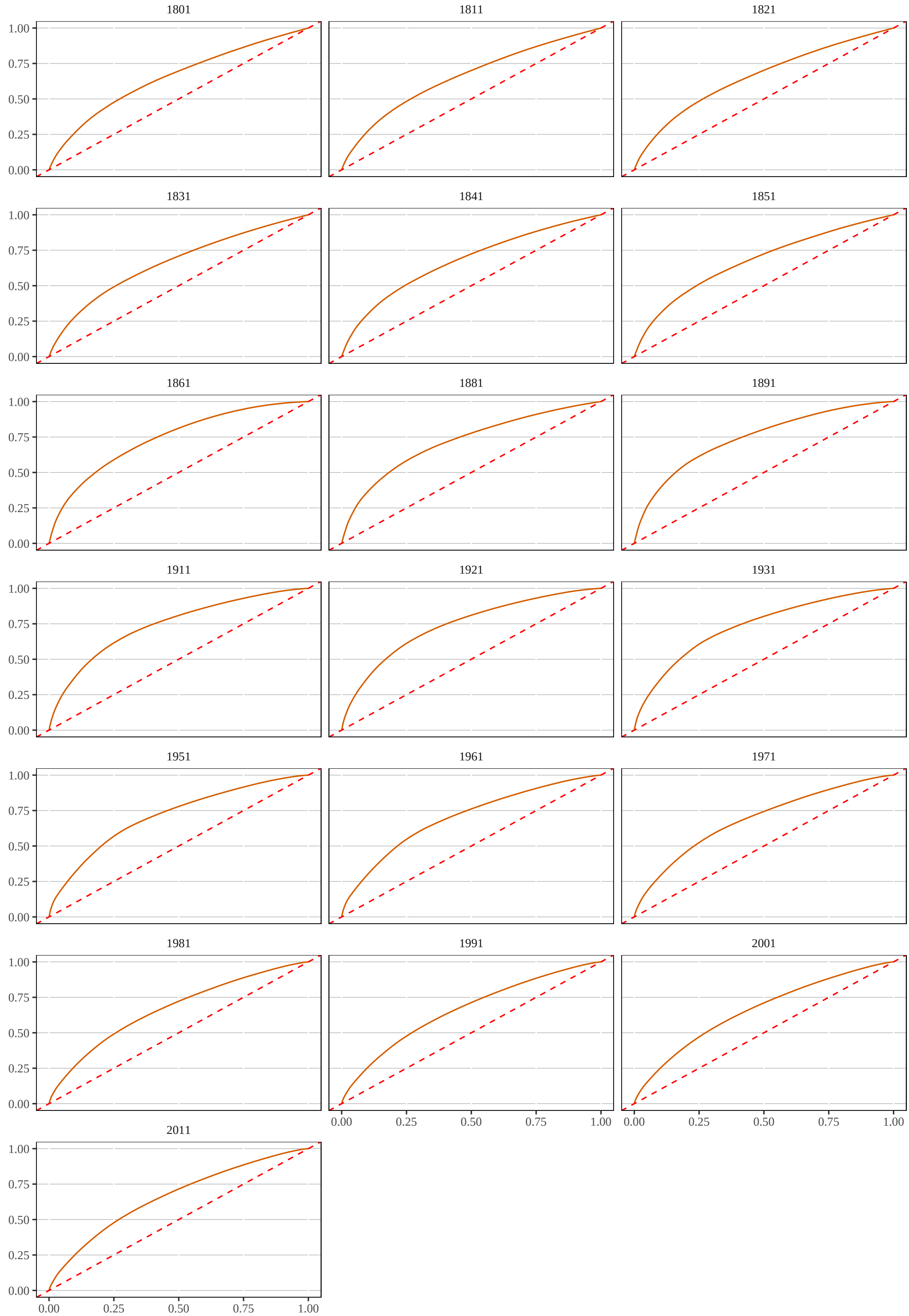


Figure 25: Lorenz curve for Fraction cutoff - Unitary Authority

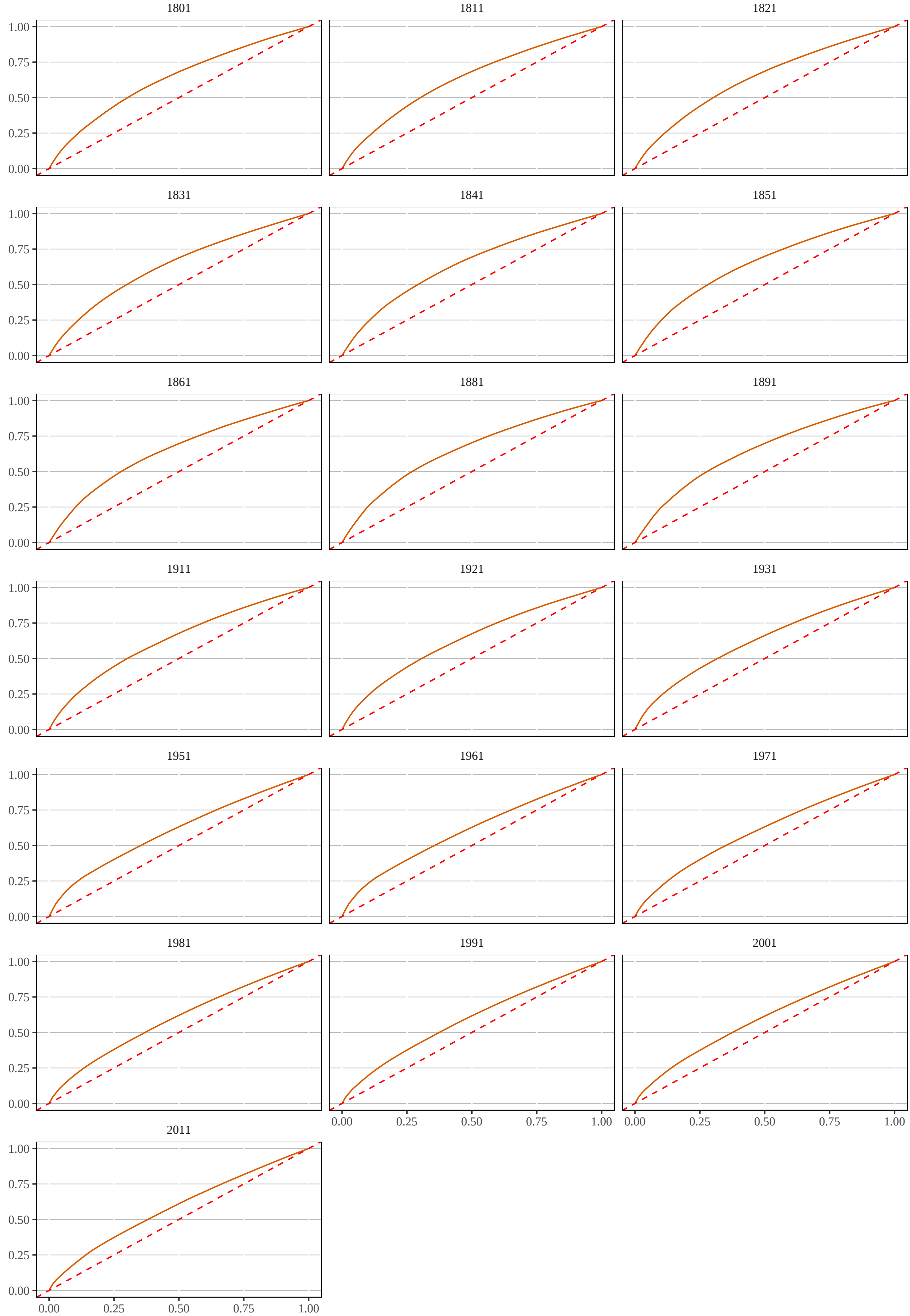


Figure 26: Lorenz curve for Level cutoff - Unitary Authority

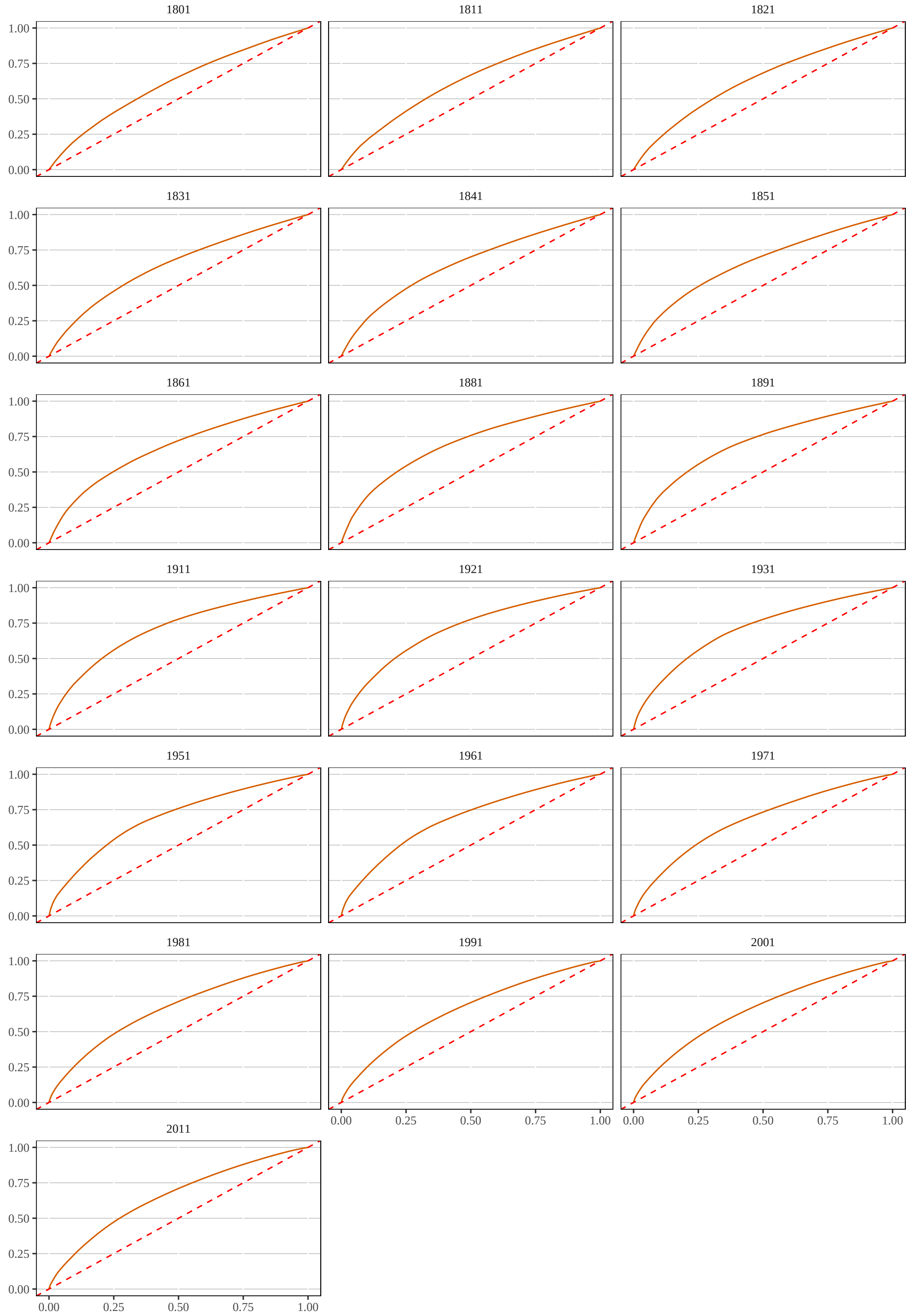
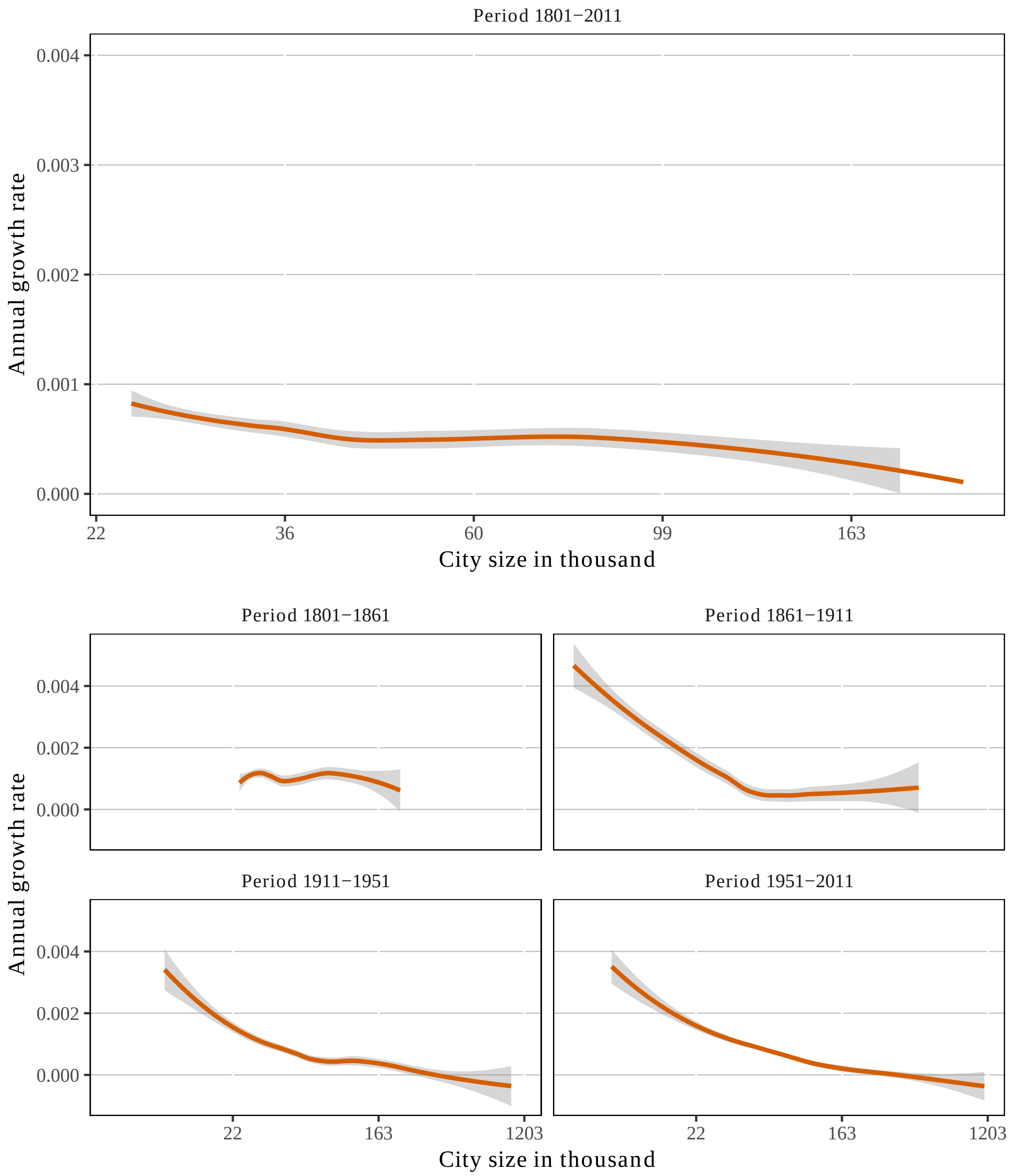


Figure 27: Kernel regression: Unitary Authority with Conservative cutoff



Note: Kernel regression: growth rate for the whole period 1801 to 2011 plotted against initial city size in 1801, and intermediate periods 1801-1861, 1861-191 1911-1951 and 1951-2011 plotted against initial size.