

Document de treball de l'IEB 2012/18

NEW EVIDENCE ON GIBRAT'S LAW FOR CITIES

Rafael González-Val, Luis Lanaspa, Fernando Sanz

Cities and Innovation

Document de
treball de l'IEB

NEW EVIDENCE ON GIBRAT'S LAW FOR CITIES

Rafael González-Val, Luis Lanaspá, Fernando Sanz

The **IEB** research program in **Cities and Innovation** aims at promoting research in the Economics of Cities and Regions. The main objective of this program is to contribute to a better understanding of agglomeration economies and 'knowledge spillovers'. The effects of agglomeration economies and 'knowledge spillovers' on the Location of economic Activities, Innovation, the Labor Market and the Role of Universities in the transfer of Knowledge and Human Capital are particularly relevant to the program. The effects of Public Policy on the Economics of Cities are also considered to be of interest. This program puts special emphasis on applied research and on work that sheds light on policy-design issues. Research that is particularly policy-relevant from a Spanish perspective is given special consideration. Disseminating research findings to a broader audience is also an aim of the program. The program enjoys the support from the **IEB-Foundation**

The **Barcelona Institute of Economics (IEB)** is a research centre at the University of Barcelona which specializes in the field of applied economics. Through the **IEB-Foundation**, several private institutions (Applus, Abertis, Ajuntament de Barcelona, Diputació de Barcelona, Gas Natural and La Caixa) support several research programs.

Postal Address:

Institut d'Economia de Barcelona
Facultat d'Economia i Empresa
Universitat de Barcelona
C/ Tinent Coronel Valenzuela, 1-11
(08034) Barcelona, Spain
Tel.: + 34 93 403 46 46
Fax: + 34 93 403 98 32
ieb@ub.edu
<http://www.ieb.ub.edu>

The IEB working papers represent ongoing research that is circulated to encourage discussion and has not undergone a peer review process. Any opinions expressed here are those of the author(s) and not those of IEB.

NEW EVIDENCE ON GIBRAT'S LAW FOR CITIES

Rafael González-Val, Luis Lanaspa, Fernando Sanz

ABSTRACT: The aim of this work is to test empirically the validity of Gibrat's law on the growth of cities, using data on the complete distribution of cities from three countries (the US, Spain and Italy) for the entire twentieth century. In order to achieve this, we use different techniques. First, panel data unit root tests tend to confirm the validity of Gibrat's law in the upper-tail distribution. Second, when we consider the entire distribution, we find that Gibrat's law does not hold exactly in the long term using nonparametric methods that relate the growth rate to the initial city size.

JEL Codes: R00, C14

Keywords: Gibrat's law, city size distribution, urban growth

Rafael González-Val
Universitat de Barcelona & IEB
Avda. Diagonal 690
08034 Barcelona, Spain
E-mail: r.gonzalez-v@ub.edu

Luis Lanaspa
Dpto. de Análisis Económico
Universidad de Zaragoza
Facultad de CC. Económicas y
Empresariales
Gran Vía, 2
50005 Zaragoza, Spain
E-mail : llanas@unizar.es

Fernando Sanz
Dpto. de Análisis Económico
Universidad de Zaragoza
Facultad de CC. Económicas
y Empresariales
Gran Vía, 2
50005 Zaragoza, Spain
E-mail : fsanz@unizar.es

1. Introduction

The relationship between the growth rate of a quantifiable phenomenon and its initial size is a question with a long history in statistics: do larger entities grow more quickly or more slowly? By contrast, perhaps no relationship exists and the rate is independent of size. A fundamental contribution to this debate is that of Gibrat (1931), who observed that the distribution of size (measured by sales or the number of employees) of firms could be approximated well with a lognormal, and that the explanation lies in the growth process of firms tending to be multiplicative and independent of their sizes. This proposition became known as Gibrat's law and prompted a deluge of work exploring the validity of this law in relation to the distribution of firms (see the surveys by Sutton (1997) and Santarelli et al. (2006)). Gibrat's law states that no regular behaviour of any kind can be deduced between growth rate and initial size. The fulfilment of this empirical proposition also has consequences for the distribution of the variable; in the words of Gibrat (1931): "*The Law of proportionate effect will therefore imply that the logarithms of the variable will be distributed following the (normal distribution).*"

In the field of urban economics, Gibrat's law, especially since the 1990s, has given rise to numerous empirical studies testing its validity for city size distributions, arriving at a majority consensus, although not absolute at all, that it explains the growth of cities relatively well and tends to hold in the long term. This has yielded theoretical works explaining the fulfilment of Gibrat's law in the context of external urban local effects and productive shocks, associating it directly with an equilibrium situation. These theoretical models include those of Gabaix (1999), Duranton (2006, 2007) and Córdoba (2008).

Returning to the empirical side, there is an apparent contradiction in these studies, because they usually accept the fulfilment of Gibrat's law but at the same time claim that the distribution followed by city size (at least the upper tail) is a Pareto distribution, which is very different from the lognormal. Eeckhout (2004) was able to reconcile both results by demonstrating (as Parr and Suzuki (1973) claimed in a pioneering work) that, imposing size restrictions on the cities, taking only the upper tail, this biases the analysis. Thus, if all cities are taken, it can be found that the true distribution is lognormal, and that the growth of these cities is independent of size. However, to date, the studies by Eeckhout (2004) and Giesen et al. (2010) are the only ones to consider the entire city size distribution. Nevertheless, these are short-term analyses,¹ and the phenomenon under study (Gibrat's law) is a long-term result (Gabaix and Ioannides, 2004).

The aim of this work is to test empirically the validity of Gibrat's law on the growth of cities, using data on the complete distribution of cities (without size restrictions) in three countries (the US, Spain and Italy) for the entire twentieth century. Our results qualify this study as the most comprehensive empirical examination of Gibrat's law to date, considering the number of cities (un-truncated settlement size data from three countries), the long time span (a whole century) and the different methodologies that we apply (parametric and nonparametric techniques).

Furthermore, the three countries selected give us information about two different urban behaviours. The US is an extremely interesting country with which to analyse the

¹ Eeckhout (2004) took data from the United States censuses of 1990 and 2000, possibly because they are the only ones to be available online. Levy (2009), in a comment to Eeckhout (2004), and Eeckhout (2009) in the reply, also considered no truncation point, but only for the 2000 US census data. Giesen et al. (2010), based on the pioneering work of Reed (2002), fitted the double Pareto lognormal (DPLN) and lognormal distributions to data from eight countries (Germany, France, the Czech Republic, Hungary, Italy, Switzerland, Brazil and the US). The data for the US were the same as those used by Eeckhout (2004, 2009) and Levy (2009). It should be noted that the DPLN also builds on Gibrat's law, particularly a generalized version of the law (Reed, 2002).

evolution of urban structure because it is a relatively young country whose inhabitants are characterised by high mobility. By contrast, the European countries have a much older urban structure and their inhabitants present greater resistance to movement; specifically, Cheshire and Magrini (2006) estimate mobility in the US to be 15 times higher than that in Europe. Spain and Italy have a consolidated urban structure and new cities rarely appear (the foundation of many European cities dates back to the Middle Ages, so urban growth is produced by population increases in existing cities). In the US, however, urban growth has a double dimension: as well as increases in city size, the number of cities also increases, with potentially different effects on city size distribution.

The following section offers a brief overview of the literature on Gibrat's law for cities and the results obtained. Section 3 introduces the databases. From our results, we deduce that panel data unit root tests tend to confirm the validity of Gibrat's law in the upper-tail distribution (section 4.1). In section 4.2, we consider the entire distribution and, using nonparametric methods, we find that Gibrat's law does not hold exactly in the long term (in general, size affects the variance of the growth process but not its mean). The validity of the law in the short term (by decade) is even weaker. In section 5, we test whether the lognormal distribution is a good description of city size distributions across the entire century. The work ends with our conclusions.

2. Gibrat's law for cities: an overview of the literature

In the 1990s, numerous studies began to appear that empirically tested the validity of Gibrat's law. Table 1 shows the classification of all the studies on urban economics that we know of to date. The countries considered, statistical and econometric techniques used and sample sizes are heterogeneous, although Table 1 shows that most of the recent empirical works have chosen between two techniques to

test for Gibrat's law: panel unit root tests to consider a small sample size and nonparametric kernel regressions when the sample size is large. We apply these two methodologies in section 4. The results are fairly mixed, with the acceptance of the law the predominant outcome, albeit by a slight margin.

Both Eaton and Eckstein (1997) and Davis and Weinstein (2002) accept its fulfilment for Japanese cities, although they use different sample sections (40 and 303 cities, respectively) and time horizons. Brakman et al. (2004) come to the same conclusion when analysing the impact of bombardment on Germany during the Second World War, concluding that, for the sample of 103 cities examined, bombing had a significant but temporary impact on post-war city growth. Nevertheless, nearly the same authors in Bosker et al. (2008) obtain a mixed result with a sample of 62 cities in West Germany: correcting for the impact of the Second World War, Gibrat's law is found to hold only for about 25% of the sample. Giesen and Südekum's (2011) recent work cannot formally reject Gibrat's law for the 71 largest German cities in 1997.

Meanwhile, both Clark and Stabler (1991) and Resende (2004) also accept the hypothesis of proportionate urban growth for Canada and Brazil, respectively. The sample size used by Clark and Stabler (1991) is tiny (the seven most populous Canadian cities), although the main contribution of their work is to propose the use of data panel methodology and unit root tests in the analysis of urban growth. Resende (2004) applies the same methodology to his sample of 497 Brazilian cities. However, Henderson and Wang (2007) strongly reject Gibrat's law and a unit root process in their worldwide data set on all metro areas over 100,000 inhabitants from 1960 to 2000.

For the case of the US, there are also several works that statistically accept the fulfilment of Gibrat's law, whether at the level of cities—Eeckhout (2004) is the first to use the entire sample without size restrictions and González-Val generalises this

analysis for the entire twentieth century—or with Metropolitan Statistical Areas (MSAs) (Ioannides and Overman (2003), whose results reproduce Gabaix and Ioannides (2004)). Also for the US, however, Black and Henderson (2003) reject Gibrat's law for any sample section, although their database of MSAs differs² from that used by Ioannides and Overman (2003). Michaels et al. (2012) use data from Minor Civil Divisions and counties to track the evolution of populations across both rural and urban areas in the United States from 1880 to 2000, finding that Gibrat's law is a reasonable approximation for population growth only for the largest units.

Other works exist that reject the fulfilment of Gibrat's law. Thus, Guérin-Pace (1995) finds that in France, using a wide sample of cities with over 2,000 inhabitants during the period 1836–1990, there seems to be a fairly strong correlation between city size and growth rate, a correlation that is accentuated when the logarithm of the population is considered. This result opposes that obtained by Eaton and Eckstein (1997) when considering only the 39 most populated French cities. Petrakos et al. (2000) and Soo (2007) also reject the fulfilment of Gibrat's law in Greece and Malaysia, respectively. For the case of China, Anderson and Ge (2005) obtain a mixed result with a sample of 149 cities of more than 100,000 inhabitants: Gibrat's law seems to describe well the situation prior to the Economic Reform and One Child Policy period, but later Kalecki's reformulation seems to be more appropriate.

What we wish to emphasize is that, with the exception of Eeckhout (2004), Giesen et al. (2010) and González-Val (2010), none of these studies considers the entire distribution of cities, because all of them impose a truncation point, whether explicitly by taking cities above a minimum population threshold or implicitly by working with

² The standard definitions of metropolitan areas were first published in 1949. This means that if the objective is to make a long-term analysis, it will be necessary to reconstruct the areas for earlier periods, in the absence of a single criterion.

MSAs.³ This is usually because of practical reasons regarding data availability. Consequently, most studies focus on analysing the most populous cities, the upper-tail distribution. However, any analysis using this kind of sample will have a local character because the behaviour of large cities cannot be extrapolated to the entire distribution. This type of deduction can lead to biased conclusions, because it must not be forgotten that what is being analysed is the behaviour of a few cities, which in addition to being of a similar size can present common patterns of growth. Therefore, we might conclude that Gibrat's law holds when in fact we have focused our analysis on a group of cities that cannot be representative of all urban centres.

3. Databases

We use un-truncated settlement size data from three countries: the US, Spain and Italy.⁴ Our database includes decennial census data for each decade of the twentieth century.⁵ Table 2 shows the number of cities for each decade and the descriptive statistics.

The data for the US are the same as those used by González-Val (2010). Our base, created from the original documents of the annual census published by the US Census Bureau, www.census.gov, consists of the available data of all incorporated places without any size restriction for each decade of the twentieth century. The US Census Bureau uses the generic term *incorporated place* to refer to a governmental unit incorporated under state Law as a city, town, borough or village, which has legally

³ In the US, to qualify as an MSA a city needs to have 50,000 or more inhabitants, or the presence of an urbanised area of at least 50,000 inhabitants, and a total metropolitan population of at least 100,000. In other countries, similar criteria are followed, although the minimum population threshold required to be considered a metropolitan area may change.

⁴ We use data from "legal" cities. However, there are problems involving international comparability because the administrative definition of a city changes from one country to another. Nevertheless, the concepts of municipality used in Spain and Italy are similar.

⁵ No census exists in Italy for 1941 because of its participation in the Second World War, so we have used the data for 1936.

established limits, powers and functions. Incorporated places in Alaska, Hawaii and Puerto Rico are excluded because of data limitations.

The alternative would be the use of metropolitan areas' data. Both units have advantages. As Glaeser and Shapiro (2003) indicate, metro areas represent urban agglomerations, covering huge areas that are meant to capture labour markets. Metropolitan areas are attractive because they are more natural economic units. Legal cities are political units that usually lie within metropolitan areas, and their boundaries make no economic sense, although some factors, such as human capital spillovers, are thought to operate at a very local level. Furthermore, Eeckhout (2004) argues that there are statistical reasons that justify the use of incorporated places (un-truncated data) rather than metro areas.

The percentage of the total US population that our sample of incorporated places represents can appear low compared with other studies using MSAs. The population of incorporated places increases from representing less than half the total population of the US in 1900 (46.99%) to 61.49% in 2000, while the number of cities increases by 82.11% from 10,596 in 1900 to 19,296 in 2000. However, it is similar to that of other works using cities.⁶ The population excluded from the sample is what the US Census Bureau calls population not in place. Incorporated places do not cover the whole territory of the US. Some territories are excluded from any recognised place. For example, more than 74 million people (26.64% of the total US population) lived in a territory that, at least officially, was not in a place in 2000.⁷ Most of these people (61.58% in 2000) are rural population.

⁶ For example, see Kim (2000) and Kim and Margo (2004), who define a city as an area having a population of greater than 2,500 inhabitants.

⁷ Census 2000 data on the population in places and not in places can be found in Table 9 of PHC-3 (US Summary, part 1), available online at: <http://www.census.gov/prod/cen2000/index.html>.

Although these people living outside incorporated places are excluded from our sample, they are included in some MSAs because these are multi-county units and this population is counted as inhabitants of the counties. MSAs cover huge geographic areas and include a large proportion of the population living in rural areas. This explains why the percentage of the total population represented by MSAs is higher than our sample of incorporated places. However, despite the sample of incorporated places covering a lower percentage of the total US population, the population of incorporated places is almost entirely urban (94.18% in 2000) compared with 88.35% of the urban population in the MSAs.

For Spain and Italy, the geographical unit of reference is the municipality. The data come from the official statistical information services. In Italy, this is the Istituto Nazionale di Statistica (www.istat.it), and for Spain our source is the censuses by the Instituto Nacional de Estadística,⁸ (INE, www.ine.es).

Municipalities are the smallest spatial units (local governments), thus they are the administratively defined “legal” cities. The main difference between these municipalities and the incorporated places is that municipalities are the lowest spatial subdivision in Spain and Italy, so they represent the whole territory of the country. Municipalities comprise the total land area, and therefore all the population too (see Table 2). However, in the US, a large amount of land area and population is not included in any place, as noted before.

Figure 1 displays the mean growth rates for each decade weighted by city size, calculated from the gross growth rates, defined as $g_{it} = \frac{S_{it} - S_{it-1}}{S_{it-1}}$, where S_{it} is the

⁸ The official INE census has been improved in an alternative database, created by Azagra et al. (2006), reconstructing the population census for the twentieth century using territorially homogeneous criteria. We have repeated the analysis using this database and the results are not significantly different, so we chose to present the results deduced from the official data.

population of the city i in the year t . In the US, it can be observed that the first decades of the century saw strong growth rates for city sizes. However, this period of growth ended in 1920–1930. Between 1930 and 1970, the growth rates rise and fall, and then rise again in the last two decades. The two periods of lowest growth, 1930–1940 and 1970–1980, are very close to two profound economic crises (the Great Depression and the second oil supply shock in 1979). Spain and Italy present lower growth rates; since the overall population did not fall in these countries in any decade, the declines in growth rates are related to composition effects: many small municipalities tend to shrink while the few large ones grow, in such a way that the weighted average decreases. The decline in growth rates at the end of the century in Spain and Italy is a consequence of the suburbanization process that took place in both European countries.

4. Testing for Gibrat's law

4.1. Parametric analysis: panel unit root testing

Clark and Stabler (1991) suggest that testing for Gibrat's law is equivalent to testing for the presence of a unit root. This idea is also emphasized by Gabaix and Ioannides (2004), who expect “*that the next generation of city evolution empirics could draw from the sophisticated econometric literature on unit root.*”. In line with this suggestion, most studies now apply unit root tests (see Table 1).

Some authors (Black and Henderson, 2003; Henderson and Wang, 2007; Soo, 2007) test the presence of a unit root by proposing a growth equation, which they estimate using panel data. Nevertheless, as pointed out by Gabaix and Ioannides (2004) and Bosker et al. (2008), this methodology presents some drawbacks. First, the periodicity of our data is by decades, and we have only 11 temporal observations (decade-by-decade city sizes over a total period of 100 years), when the ideal would be

to have at least annual data. Second, the presence of cross-sectional dependence across the cities in the panel can give rise to estimations that are not very robust. It has been well established in the literature that panel unit root and stationarity tests that do not explicitly allow for this feature among individuals present size distortions (Banerjee et al., 2005).

For this, we use one of the tests especially created to deal with this question: Pesaran's (2007) test for unit roots in heterogeneous panels with cross-section dependence is calculated based on the CADF statistic (cross-sectional ADF (see below) statistic). To eliminate cross-dependence, the standard Dickey–Fuller (or Augmented Dickey–Fuller (ADF)) regressions are augmented with the cross-section averages of lagged levels and first-differences of the individual series, such that the influence of the unobservable common factor is asymptotically filtered.

The test of the unit root hypothesis is based on the t-ratio of the OLS estimate of b_i in the following cross-sectional augmented DF (CADF) regression:

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + d_i \Delta \bar{y}_t + e_{it}, \quad (1)$$

where a_i is the individual city-specific average growth rate and \bar{y}_t is the cross-section mean of y_{it} , $\bar{y}_t = N^{-1} \sum_{j=1}^N y_{jt}$. We will test for the presence of a unit root in the natural logarithm of city relative size ($y_{it} = \ln s_{it}$). City relative size (s_{it}) is defined as

$$s_{it} = \frac{S_{it}}{\bar{S}_t} = \frac{S_{it}}{\frac{1}{N} \sum_{i=1}^N S_{it}}; \text{ from a long-term temporal perspective of steady-state}$$

distributions, it is necessary to use a relative measure of size (Gabaix and Ioannides, 2004). The null hypothesis assumes that all series are nonstationary, and Pesaran's CADF is consistent under the alternative that only a fraction of the series is stationary.

However, the problem with Pesaran's test is that it is not designed to deal with such large panels (22,078 cities in the US, 8,077 in Spain and 8,100 in Italy), especially when so few temporal observations are available ($N \rightarrow \infty, T = 11$). For this reason, we must limit our analysis to the largest cities (although the next section offers a nonparametric analysis of the entire sample).⁹

As noted previously in the literature review in section 2, traditionally most studies focus only on the upper-tail distribution, because of the data availability. Recent papers (Eeckhout, 2009; Levy, 2009) discuss that the behaviour of the upper-tail distribution can differ from that of the rest of the sample, because for the largest cities the lognormal tail and the Pareto tail of the distribution are hard to distinguish. Moreover, as Levy (2009) argues, while the upper tail in the US city size distribution in 2000 includes only 6% of the cities, it accounts for almost 23% of the total US population. Therefore, a separate analysis for the largest cities is important.

Table 3 shows the results of Pesaran's test, both the value of the test statistic and the corresponding p-value, applied to the upper-tail distribution until the 500 largest cities in the initial period have been considered for all the decades. All the statistics are based on univariate AR(1) specifications including constant and trend. The null hypothesis of a unit root is not rejected in the US or Italy for any of the sample sizes considered, providing evidence in favour of the long-term validity of Gibrat's law. Spain's case is different, since when the sample size is more than the 200 largest cities the unit root is rejected. The analysis in the next section reveals that the reason for this

⁹ Pesaran's test performs well for a small T and large N , but the power of the test decreases dramatically when, as in our case, N is large and T is small. The Monte Carlo results in Pesaran (2007) are in line with the theoretical findings of Moon et al. (2007), who show that local power of panel unit root tests is in the $N^{-1/4}T^{-1}$ neighbourhood of the null in the case of models with linear trends. In our case, with $N \rightarrow \infty$ and $T = 11$, the power of the test would be almost zero if all cities were considered.

rejection is a positive relationship between the relative size and the growth rate for the largest cities. This result could be a consequence of the political regime, a military dictatorship in most decades of the century. In this context, although only for the capital city, Ades and Glaeser (1995) find that this city will tend to be more dominant the more political instability there is in a country and the more authoritarian is its regime.

4.2. Nonparametric analysis: kernel regression conditional on city size

Here we perform an analysis of the entire distribution, not just the upper tail. As a first approximation to city growth, Figure 2 shows the scatter plots of growth against city relative size for three representative decades (the behaviour in the rest of the adjacent periods is similar) in the US, Spain and Italy. These graphs seem to support that growth is independent of size, although they also point to a great variance across observations, especially in the case of the US. In this section, we analyse the relationship between growth and initial size using two different nonparametric tools.

First, we perform a nonparametric analysis using kernel regressions as in Ioannides and Overman (2003) and Eeckhout (2004). It consists of taking the following specification:

$$g_i = m(s_i) + \varepsilon_i, \quad (2)$$

where g_i is the growth rate $(\ln s_{it} - \ln s_{it-1})$ normalised (subtracting the contemporary mean and dividing by the standard deviation in the relevant decade) and s_i is the logarithm of the i th city's relative size. Instead of making assumptions about the functional relationship m , $\hat{m}(s)$ is estimated as a local mean around the point s and is smoothed using a kernel, which is a symmetrical, weighted and continuous function in s .

To estimate $\hat{m}(s)$, the Nadaraya–Watson method is used, exactly as it appears in Härdle (1990, Chapter 3), based on the following expression:¹⁰

$$\hat{m}(s) = \frac{n^{-1} \sum_{i=1}^n K_h(s - s_i) g_i}{n^{-1} \sum_{i=1}^n K_h(s - s_i)}, \quad (3)$$

where K_h denotes the dependence of the kernel K (in this case an Epanechnikov) on the bandwidth h . We use the same bandwidth (0.5) in all the estimations to permit comparisons between countries.

Starting from this calculated mean $\hat{m}(s)$, the variance of the growth rate g_i is also estimated, again by applying the Nadaraya–Watson estimator:

$$\hat{\sigma}^2(s) = \frac{n^{-1} \sum_{i=1}^n K_h(s - s_i) (g_i - \hat{m}(s))^2}{n^{-1} \sum_{i=1}^n K_h(s - s_i)}. \quad (4)$$

The estimator is very sensitive, both in mean and in variance, to atypical values. Eeckhout (2004) finds that some outliers have a huge impact on the variance of growth rates. For this reason, in the same way as Eeckhout (2004), we eliminate from the sample some atypical observations: the 5% of smallest cities because they usually have much higher growth rates in mean and variance. This is logical; these are cities of under 200 inhabitants, where the smallest increase in population is very large in percentage terms.

Following Gabaix and Ioannides (2004), “*Gibrat’s law states that the growth rate of an economic entity (firm, mutual fund, city) of size S has a distribution function*

¹⁰ The calculation was performed with the KERNREG2 Stata module, developed by Cox, Salgado-Ugarte, Shimizu and Taniuchi, and available online at: <http://ideas.repec.org/c/boc/bocode/s372601.html>.

with mean and variance that are independent of S ". Thus, they distinguish between Gibrat's law for means and Gibrat's law for variances. As the growth rates are normalised, if Gibrat's law in mean is strictly fulfilled, the nonparametric estimate will be a straight line on the zero value. Values different from zero involve deviations from the mean. Moreover, the estimated variance of the growth rate will also be a straight line on the value one, which would mean that the variance does not depend on the size of the city. To be able to test these hypotheses, we construct bootstrapped 95% confidence bands (calculated from 500 random samples with replacements).

We offer a first approach to the behaviour of city growth from a short-term perspective, i.e., considering each decade individually. Figures 3, 4 and 5 show the nonparametric estimates for the US, Spain and Italy, respectively, corresponding to the same three representative decades shown in Figure 2.

Two different behaviours can be observed; while in the US the estimate of growth is very close to the zero value (this value falls within the confidence bands for most of the distributions, supporting Gibrat's law even in the short term), in Spain and Italy a different pattern of growth can be seen. Starting from the beginning of the century until the mid-century, the city growth exhibits clear divergent behaviour in both European countries, although Gibrat's law can only be rejected for some values at the upper-tail distribution. However, in the second half of the century the growth changes gradually to an inverted U-shaped pattern. These results confirm that, as Gabaix and Ioannides (2004) indicate, *"the casual impression of the authors is that in some decades, large cities grow faster than small cities, but in other decades, small cities grow faster"*.

There is a negative relationship between the estimated variance of growth and city size in the three countries for most of the decades (this is especially true for the US,

where Gibrat's law can be rejected at the upper tail), although in Spain and Italy the behaviour of the variance is irregular, particularly in the first decades of the century.

Moreover, to analyse the entire twentieth century, we build a pool with all the growth rates between two consecutive periods. This enables us to carry out long-term analysis. Figure 6 shows the nonparametric estimates of the growth rate of a pool for the entire twentieth century for the US (1900–2000, 152,475 observations), Spain (1900–2001, 74,100 observations) and Italy (1901–2001, 73,260 observations). For the US, the value zero is always in the confidence bands, so that the growth rates being significantly different for any city size cannot be rejected. For Spain and Italy, the estimated mean grows with the sample size, although it is significantly different from zero only for the largest cities. One possible explanation is historical: both Spain and Italy suffered wars on their territories during the twentieth century, so that for several decades the largest cities attracted most of the population.¹¹ However, the estimations by decade indicate that this tendency would have reversed in the second half of the century. Therefore, we find evidence in favour of Gibrat's law for means for the US throughout the twentieth century. Support is weaker in Spain and Italy because the largest cities show some divergent behaviour.

Figure 6 also shows the nonparametric estimates of the variance of growth rates of a pool for the entire twentieth century for the US, Spain and Italy. As expected, while for most of the distribution the value one falls within the confidence bands, indicating that there are no significant differences in variance, the tails of the distribution show differentiated behaviours. In the US, the variance clearly decreases with the size of the city, while in Spain and Italy the behaviour is more erratic and the biggest cities also have high variances.

¹¹ This result could be related to the “safe harbour effect” of Glaeser and Shapiro (2002), which is a centripetal force that tends to agglomerate the populations in large cities when there is an armed conflict.

Our results, obtained with a sample of all incorporated places without any size restriction, are similar to those obtained by Ioannides and Overman (2003), with their database of MSAs. To sum up, the nonparametric estimates (Figure 6) show that while the mean of growth (Gibrat's law for means) seems to be independent of size in the three countries in the long term (although in Spain and Italy the largest cities present some divergent behaviour), the variance of growth (Gibrat's law for variances) depends negatively on size: the smallest cities present clearly higher variances in all three countries (although in Spain and Italy the behaviour is more erratic). In the short term (Figures 3, 4 and 5), the evidence supporting Gibrat's law is weaker, as it corresponds to a law that is thought to hold mainly in the long term (Gabaix and Ioannides, 2004). This points to Gibrat's law holding weakly (growth is proportional in means but not in variance).¹²

Finally, as González-Val (2012), we perform a nonparametric estimation of growth using a resistant smoothing approach. Kernel estimation of regression functions has been receiving a great deal of attention in the recent literature examining Gibrat's law (Ioannides and Overman, 2003; Eeckhout, 2004; González-Val, 2010; Giesen and Südekum, 2011) and the most widely used estimator is the Nadaraya–Watson estimator. Thus, the previous results can be compared with those of other studies. However, as argued before, the Nadaraya–Watson estimator is known to be highly sensitive to the presence of outliers in the data, so we have to exclude some observations.

Now, we try to reduce this sensitivity by using a resistant smoothing technique, the LOcally WEighted Scatter plot Smoothing (LOWESS) algorithm, proposed by

¹² Although our results support Gibrat's law only partially, there are theoretical models that explain how growth can be proportional in means but not in variance. Gabaix (1999) contemplates the possibility that Gibrat's law might not hold exactly, and examines the case in which cities grow randomly with expected growth rates and standard deviations that depend on their sizes. Córdoba (2008) also introduces a parsimonious generalisation of Gibrat's law that allows size to affect the variance of the growth process but not its mean.

Cleveland (1979). It is based on local polynomial fits; see Härdle (1990, Chapter 6). The advantages of LOWESS are that it is a free-functional form method¹³ and that it is robust to atypical values. Therefore, it allows us to obtain robust nonparametric estimates of growth and variance, using the entire sample, including the smallest 5% of distribution observations, which we previously excluded. Figure 7 shows the results for the twentieth-century pool of observations for the US, Spain and Italy (this figure is analogous to Figure 6, but includes the smallest 5% of observations).¹⁴ Their inclusion produces an increase in the estimates of both growth and variance at the lower tail of the distribution; this increment is much greater in the case of variance, as the dispersion of these observations is very high. Thus, small cities exhibit higher growth (except for the case of Italy) and variance than the rest of the cities, indicating again that Gibrat's law does not hold exactly.

For the rest of the sample, the results estimated by LOWESS are very similar to those estimated by the Nadaraya–Watson estimator, both in growth and in variance (Figure 6). This is logical because the Nadaraya–Watson estimator estimates a local mean around each point of the grid, so the estimates for the medium-size cities or the upper-tail distribution do not depend on the values of the smallest observations, indicating that our previous results excluding the smallest 5% of observations are robust for most of the distribution. Their inclusion only increases growth and variance at the lower tail of the distribution.

¹³ It does not require the specification of a function to fit a model to all of the data in the sample; LOWESS simply carries out a locally weighted regression of the y variable on the x variable, obtaining a new smoothed variable. We use the `lowess` command in STATA with the default options: a smoothing parameter equal to 0.8 and a tricube weighting function.

¹⁴ Estimates by decade, not shown, are available from the authors on request.

5. What about city size distribution?

Proportionate growth implies a lognormal distribution, and this is a statistical relationship (Gibrat, 1931; Kalecki, 1945). However, if there is a lower bound to the distribution (which can be very low) the resulting distribution is Pareto (Gabaix, 1999), so, as Eeckhout (2004) shows, city size distribution follows a lognormal distribution only when we consider all cities without any size restriction.¹⁵ Our results show that the growth process leads to a lognormal distribution with standard deviation that increases in time t in the three countries if all cities are considered. This is result theoretically predicted: under a Brownian motion (Gabaix, 1999; Ioannides and Overman, 2003), the sample standard deviation should increase with the passing of time as a function of t (Anderson and Ge, 2005). Furthermore, it can be shown that by introducing a small change in one of the assumptions in Kalecki's classic model,¹⁶ the same standard framework can be obtained, combining lognormality with a variance that increases in time.

We carried out Wilcoxon's lognormality test (rank-sum test), which is a nonparametric test for assessing whether two samples of observations come from the same distribution. The null hypothesis is that the two samples are drawn from a single population and, therefore, that their probability distributions are equal, in our case, the lognormal distribution. Wilcoxon's test has the advantage of being appropriate for any sample size. The more frequent normality tests – Kolmogorov–Smirnov, Shapiro–Wilks, D'Agostino–Pearson – are designed for small samples and so tend to reject the null hypothesis of normality for large sample sizes, although the deviations from lognormality are arbitrarily small.

¹⁵ We acknowledge one anonymous referee for this comment.

¹⁶ Just changing Equation (2) of Kalecki (1945), making $\sum Yy > 0$.

Table 4 shows the results of the test. The conclusion is that the null hypothesis of lognormality cannot be rejected at 5% for all the periods of the twentieth century in Spain and Italy. In the US, a temporal evolution can be observed; in the first decades, lognormality is rejected and the p-value decreases over time, but from 1930 the p-value begins to grow until lognormal distribution is not rejected at 5% from 1960 onwards. In fact, if instead of 5% we take a significance level of 1%, the null hypothesis would only be rejected in 1920 and 1930.

However, the shape of the distribution in the US for the period 1900–1950 is not far from lognormality either. Figure 8 shows the empirical density functions estimated by adaptive Gaussian kernels for 1900, 1950 (the last year in which lognormality is rejected at 5%) and 2000. The motive for this systematic rejection seems to be an excessive concentration of density in the central values that is higher than would correspond to the theoretical lognormal distribution (dotted line). Starting in 1900 with a very leptokurtic distribution with a great deal of density concentrated in the mean value, from 1930 (not shown), when the growth of the urban population slows, the distribution loses kurtosis and the concentration decreases, not rejecting lognormality statistically at 5% from 1960.

To sum up, both the test carried out and the visualisation of the estimated empirical density functions seem to corroborate that city size distribution can be approximated correctly as lognormal (in Spain and Italy for the entire twentieth century and in the US for most decades, depending on the significance level).

6. Conclusions

The aim of this work is simple: to provide additional information on whether Gibrat's law, an empirical regularity that is well-known in the literature on urban

economics, holds. Briefly, this law states that the population growth rate of cities is a process deriving from independent multiplicative shocks, which implies two statistical conclusions. First, the empirical city size distribution can be well fitted by a lognormal; second, the growth rate is on average independent of the initial size of the urban centres and its evolution is fundamentally stochastic without any fixed pattern of behaviour. Moreover, although this issue is not dealt with here, if the urban growth process follows Gibrat's law, this has some theoretical implications (see the survey by Gabaix and Ioannides, 2004).

This article contributes in several ways. On the one hand, it uses a database covering un-truncated settlement size data from three countries (the US, Spain and Italy) with different urban histories, over a long time span (the entire twentieth century). As far as we know, this is the widest-ranging attempt to test the geographical and temporal validity of this law, focusing on robust results. On the other hand, it employs different methods (parametric and nonparametric).

There are two basic conclusions. First, the panel data unit root tests carried out confirm that, in the long term, Gibrat's law always holds for the upper-tail of the distribution for the US and Italy, and only for the 200 largest cities for Spain. In any case, the use of panel techniques for three countries and eleven census periods is innovative. However, from the use of nonparametric techniques considering all the cities, also over the long term, such as kernel regressions conditional on city size, we deduce that Gibrat's law does not hold exactly for the whole distribution. Gibrat's law for means seems to hold for the US and, to a lesser extent, for Spain and Italy. In these two European countries, there is a positive relationship between city size and growth, although this divergent behaviour is only significant for the largest cities. Nevertheless, we also find that, in general, the variances depend negatively on size, pointing to a weak

version of Gibrat's law where growth is proportional in means but not in variance. Moreover, small cities clearly exhibit higher growth (except for the case of Italy) and variance than the rest of the cities, even when we estimate using a nonparametric resistant smoothing approach. In the short term, as could be anticipated, the evidence regarding the validity of the law is more mixed.

Second, the lognormal distribution works well as a description of empirical city size distributions across the entire century when no truncation point is considered. Wilcoxon's rank-sum test shows that, except for the US in the first half of the century, the lognormal distribution can never be rejected.

Acknowledgements

The authors acknowledge financial support from the Spanish Ministerio de Educación y Ciencia (ECO2009-09332 and ECO2010-16934 projects), the DGA (ADETRE research group), FEDER, and the Generalitat (2009SGR102). Earlier versions of this paper were presented at the 55th North American Meeting of the Regional Science Association International (New York, 2008), at the XXXIII Symposium of Economic Analysis (Zaragoza, 2008), at the 67th International Atlantic Economic Conference (Rome, 2009), at the 24th Annual Congress of the European Economic Association (Barcelona, 2009), at the XXXV Reunión de Estudios Regionales (Valencia, 2009), at the 9th Annual Conference of the European Economics and Finance Society (Athens, 2010) and at the 50th Congress of the European Regional Science Association (Jönköping, 2010), with all the comments made by participants being highly appreciated. All remaining errors are ours.

References

- [1] Ades, A. F. and Glaeser, E. L. (1995). Trade and circuses: explaining urban giants. *The Quarterly Journal of Economics*, 110: 195–227.
- [2] Anderson, G. and Ge, Y. (2005). The Size Distribution of Chinese Cities. *Regional Science and Urban Economics*, 35: 756–776.

- [3] Azagra, J., Chorén, P., Goerlich, F. J. and Mas, M. (2006). La localización de la población española sobre el territorio. Un siglo de cambios: un estudio basado en series homogéneas (1900–2001). Fundación BBVA.
- [4] Banerjee, A., Massimiliano M. and Osbat, C. (2005). Testing for PPP: should we use panel methods? *Empirical Economics*, 30: 77–91.
- [5] Black, D. and Henderson, V. (2003). Urban evolution in the USA. *Journal of Economic Geography*, 3(4): 343–372.
- [6] Bosker, M., Brakman S., Garretsen H. and Schramm, M. (2008). A century of shocks: the evolution of the German city size distribution 1925–1999. *Regional Science and Urban Economics* 38: 330–347.
- [7] Brakman, S., Garretsen, H. and Schramm, M. (2004). The Strategic Bombing of German Cities during World War II and its Impact on City Growth. *Journal of Economic Geography*, 4: 201–218.
- [8] Cheshire, P. C. and Magrini, S. (2006). Population growth in European cities: weather matters - but only nationally. *Regional Studies*, 40(1): 23–37.
- [9] Clark, J. S. and Stabler, J. C. (1991). Gibrat's Law and the Growth of Canadian Cities. *Urban Studies*, 28(4): 635–639.
- [10] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74: 829–836.
- [11] Córdoba, J. C. (2008). A generalized Gibrat's Law for Cities. *International Economic Review*, 49(4): 1463–1468.
- [12] Davis, D. R. and Weinstein D. E. (2002). Bones, bombs, and break points: the geography of economic activity. *American Economic Review*, 92(5): 1269–1289.

- [13] Duranton, G. (2006). Some Foundations for Zipf's Law: product proliferation and local spillovers. *Regional Science and Urban Economics*, 36: 542–563.
- [14] Duranton, G. (2007). Urban evolutions: the fast, the slow, and the still. *American Economic Review*, 97(1): 197–221.
- [15] Eaton, J. and Eckstein, Z. (1997). Cities and growth: theory and evidence from France and Japan. *Regional Science and Urban Economics*, 27(4–5): 443–474.
- [16] Eeckhout, J. (2004). Gibrat's Law for (All) Cities. *American Economic Review*, 94(5): 1429–1451.
- [17] Eeckhout, J. (2009). Gibrat's Law for (all) Cities: Reply. *American Economic Review*, 99(4): 1676–1683.
- [18] Gabaix, X. (1999). Zipf's Law for Cities: An Explanation. *Quarterly Journal of Economics*, 114(3): 739–767.
- [19] Gabaix, X. and Ioannides, Y. M. (2004). The Evolution of City Size Distributions. In *Handbook of Urban and Regional Economics*, vol. 4, ed. John V. Henderson and Jean. F. Thisse, 2341–2378. Amsterdam: Elsevier Science, North-Holland.
- [20] Gibrat, R. (1931). *Les Inégalités Économiques*. Paris: Librairie du Recueil Sirey.
- [21] Giesen K. and Südekum J. (2011). Zipf's Law for cities in the regions and the country. *Journal of Economic Geography*, 11(4): 667–686.
- [22] Giesen, K., Zimmermann A. and Südekum J. (2010). The size distribution across all cities – double Pareto lognormal strikes. *Journal of Urban Economics*, 68: 129–137.
- [23] Glaeser, E. L. and Shapiro, J. M. (2002). Cities and warfare: the impact of terrorism on urban form. *Journal of Urban Economics*, 51: 205–224.

- [24] Glaeser, E. L. and Shapiro, J. M. (2003). Urban Growth in the 1990s: Is city living back? *Journal of Regional Science*, 43(1): 139–165.
- [25] González-Val, R. (2010). The evolution of the US city size distribution from a long-run perspective (1900–2000). *Journal of Regional Science*, 50(5): 952–972.
- [26] González-Val, R. (2012). A nonparametric estimation of the local Zipf exponent for all US cities. Forthcoming in *Environment and Planning B: Planning and Design*.
- [27] Guérin-Pace, F. (1995). Rank-size distribution and the process of urban growth. *Urban Studies*, 32(3): 551–562.
- [28] Härdle, W. (1990). *Applied Nonparametric Regression*. In *Econometric Society Monographs*. Cambridge, New York and Melbourne: Cambridge University Press.
- [29] Henderson, V. and Wang, H. G. (2007). Urbanization and city growth: the role of institutions. *Regional Science and Urban Economics*, 37(3): 283–313.
- [30] Ioannides, Y. M. and Overman, H. G. (2003). Zipf's Law for cities: an empirical examination. *Regional Science and Urban Economics*, 33: 127–137.
- [31] Kalecki, M. (1945). On the Gibrat Distribution. *Econometrica*, 13(2): 161–170.
- [32] Kim, S. (2000). Urban Development in the United States. *Southern Economic Journal*, 66: 855–880.
- [33] Kim, S. and Margo R. A. (2004). Historical Perspectives on U.S. Economic Geography. *Handbook of Urban and Regional Economics*, vol. 4, J. V. Henderson and J. F. Thisse, eds. Amsterdam: Elsevier Science, North-Holland, Chapter 66, pp. 2982–3019.

- [34] Levy, M. (2009). Gibrat's Law for (all) Cities: A Comment. *American Economic Review*, 99(4): 1672–1675.
- [35] Michaels, G., Rauch F. and Redding S. J. (2012). Urbanization and Structural Transformation. *The Quarterly Journal of Economics*, 127(2): 535–586 .
- [36] Moon, H.R., Perron B., and Phillips P. C. B. (2007). Incidental trends and the power of panel unit root tests. *Journal of Econometrics*, 141(2): 416–459.
- [37] Parr, J. B. and Suzuki, K. (1973). Settlement populations and the lognormal distribution. *Urban Studies*, 10: 335–352.
- [38] Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22: 265–312.
- [39] Petrakos, G., Mardakis, P. and Caraveli, H. (2000). Recent developments in the Greek system of urban centres. *Environment and Planning B: Planning and Design*, 27(2): 169–181.
- [40] Reed, W. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, 42: 1–17.
- [41] Resende, M. (2004). Gibrat's Law and the growth of cities in Brazil: a panel data investigation. *Urban Studies*, 41(8): 1537–1549.
- [42] Santarelli, E., Klomp, L. and Thurik, R. (2006). Gibrat's Law: An Overview of the Empirical Literature. In *Entrepreneurship, Growth, and Innovation: the Dynamics of Firms and Industries*, ed. Enrico Santarelli, 41–73. New York: Springer.
- [43] Soo, K. T. (2007). Zipf's Law and urban growth in Malaysia. *Urban Studies*, 44(1): 1–14.
- [44] Sutton, J. (1997). Gibrat's Legacy. *Journal of Economic Literature*. 35(1): 40–59.

Table 1. Empirical Studies on Gibrat's Law. A Survey

| Study | Country | Period | Truncation point | Sample size | GL | EcIss |
|------------------------------|---------------------------------|--------------------------------|---|--|----|------------------------------------|
| Eaton and Eckstein (1997) | France and Japan | 1876–1990 (F) 1925–1985 (J) | Cities > 50,000 inhabitants (F) Cities > 250,000 inhabitants (J) | 39 (F), 40 (J) | A | par (gr reg); non par (tr mat, lz) |
| Davis and Weinstein (2002) | Japan | 1925–1965 | Cities > 30,000 inhabitants | 303 | A | par (purt) |
| Bosker et al. (2008) | West Germany | 1925–1999 | Cities > 50,000 inhabitants | 62 | M | par (purt); non par (ker) |
| Giesen and Südekum (2011) | West Germany | 1975–1997 | Cities > 100,000 inhabitants | 71 | A | non par (ker) |
| Brakman et al. (2004) | Germany | 1946–1963 | Cities > 50,000 inhabitants | 103 | A | par (purt) |
| Clark and Stabler (1991) | Canada | 1975–1984 | Seven most populous cities | 7 | A | par (purt) |
| Resende (2004) | Brazil | 1980–2000 | Cities > 1,000 inhabitants | 497 | A | par (purt) |
| Eeckhout (2004) | US | 1990–2000 | All cities | 19,361 | A | par (gr reg); non par (ker) |
| Ioannides and Overman (2003) | US | 1900–1990 | All MSAs | 112 (1900) to 334 (1990) | A | non par (ker) |
| Gabaix and Ioannides (2004) | US | 1900–1990 | All MSAs | 112 (1900) to 334 (1990) | A | non par (ker) |
| Black and Henderson (2003) | US | 1900–1990 | All MSAs | 194 (1900) to 282 (1990) | R | par (purt) |
| González-Val (2010) | US | 1900–2000 | All cities | 10,596 to 19,296 | A | non par (ker) |
| Michaels et al. (2012) | US | 1880–2000 | Minor Civil Divisions & counties | 10,864 | M | non par (dsf) |
| Guérin-Pace (1995) | France | 1836–1990 | Cities > 2,000 inhabitants | 675 (1836) to 1782 (1990) | R | par (corr) |
| Soo (2007) | Malaysia | 1957–2000 | Urban areas > 10,000 inhabitants | 44 (1957) to 171 (2000) | R | par (purt) |
| Petrakos et al. (2000) | Greece | 1981–1991 | Urban centres > 5,000 inhabitants | 150 | R | par (gr reg) |
| Henderson and Wang (2007) | World | 1960–2000 | Metro areas > 100,000 inhabitants | 1,220 (1960) to 1,644 (2000) | R | par (purt) |
| Anderson and Ge (2005) | China | 1961–1999 | Cities > 100,000 inhabitants | 149 | M | par (rank reg); non par (tr mat) |
| Gibrat's Law: GL | EcIss: Econometric Issues | | gr reg: growth regressions | corr: coefficient of correlation (Pearson) | | |
| A: Accepted | par: parametric methods | | ker: kernel regressions | lz: Lorenz curves | | |
| R: Rejected | non par: non parametric methods | | rank reg: rank regressions | dsf: discrete-step function | | |
| M: Mixed Results | purt: panel unit root tests | | tr mat: transition matrices | | | |

Table 2. Number of Cities and Descriptive Statistics by Year and Country

| US | | | | | | | |
|-------|--------|----------|--------------------|---------|-----------|-------------------------|--------------------------------|
| Year | Cities | Mean | Standard deviation | Minimum | Maximum | Country Population (CP) | Percentage of CP in our sample |
| 1900 | 10,596 | 3,376.04 | 42,323.90 | 7 | 3,437,202 | 76,212,168 | 46.9 |
| 1910 | 14,135 | 3,560.92 | 49,351.24 | 4 | 4,766,883 | 92,228,496 | 54.6 |
| 1920 | 15,481 | 4,014.81 | 56,781.65 | 3 | 5,620,048 | 106,021,537 | 58.6 |
| 1930 | 16,475 | 4,642.02 | 67,853.65 | 1 | 6,930,446 | 123,202,624 | 62.1 |
| 1940 | 16,729 | 4,975.67 | 71,299.37 | 1 | 7,454,995 | 132,164,569 | 63.0 |
| 1950 | 17,113 | 5,613.42 | 76,064.40 | 1 | 7,891,957 | 151,325,798 | 63.5 |
| 1960 | 18,051 | 6,408.75 | 74,737.62 | 1 | 7,781,984 | 179,323,175 | 64.5 |
| 1970 | 18,488 | 7,094.29 | 75,319.59 | 3 | 7,894,862 | 203,302,031 | 64.5 |
| 1980 | 18,923 | 7,395.64 | 69,167.91 | 2 | 7,071,639 | 226,542,199 | 61.8 |
| 1990 | 19,120 | 7,977.63 | 71,873.91 | 2 | 7,322,564 | 248,709,873 | 61.3 |
| 2000 | 19,296 | 8,968.44 | 78,014.75 | 1 | 8,008,278 | 281,421,906 | 61.5 |
| Spain | | | | | | | |
| Year | Cities | Mean | Standard deviation | Minimum | Maximum | Country Population (CP) | Percentage of CP in our sample |
| 1900 | 7,800 | 2,282.40 | 10,177.75 | 78 | 539,835 | 18,616,630 | 95.6 |
| 1910 | 7,806 | 2,452.01 | 11,217.02 | 92 | 599,807 | 19,990,669 | 95.7 |
| 1920 | 7,812 | 2,621.92 | 13,501.02 | 82 | 750,896 | 21,388,551 | 95.8 |
| 1930 | 7,875 | 2,892.18 | 17,513.90 | 79 | 1,005,565 | 23,677,095 | 96.2 |
| 1940 | 7,896 | 3,180.65 | 20,099.96 | 11 | 1,088,647 | 26,014,278 | 96.5 |
| 1950 | 7,901 | 3,479.86 | 26,033.29 | 64 | 1,618,435 | 28,117,873 | 97.8 |
| 1960 | 7,910 | 3,801.71 | 33,652.11 | 51 | 2,259,931 | 30,582,936 | 98.3 |
| 1970 | 7,956 | 4,240.98 | 43,971.93 | 10 | 3,146,071 | 33,956,047 | 99.4 |
| 1981 | 8,034 | 4,701.40 | 45,995.35 | 5 | 3,188,297 | 37,742,561 | 100.0 |
| 1991 | 8,077 | 4,882.27 | 45,219.85 | 2 | 3,084,673 | 39,433,942 | 100.0 |
| 2001 | 8,077 | 5,039.37 | 43,079.46 | 7 | 2,938,723 | 40,847,371 | 99.6 |
| Italy | | | | | | | |
| Year | Cities | Mean | Standard deviation | Minimum | Maximum | Country Population (CP) | Percentage of CP in our sample |
| 1901 | 7,711 | 4,274.84 | 14,424.61 | 56 | 621,213 | 32,963,000 | 100.0 |
| 1911 | 7,711 | 4,648.11 | 17,392.98 | 58 | 751,211 | 35,842,000 | 100.0 |
| 1921 | 8,100 | 4,863.80 | 20,031.61 | 58 | 859,629 | 39,397,000 | 100.0 |
| 1931 | 8,100 | 5,067.10 | 22,559.85 | 93 | 960,660 | 41,043,000 | 100.0 |
| 1936 | 8,100 | 5,234.38 | 25,274.48 | 116 | 1,150,338 | 42,398,000 | 100.0 |
| 1951 | 8,100 | 5,866.12 | 31,137.52 | 74 | 1,651,393 | 47,516,000 | 100.0 |
| 1961 | 8,100 | 6,249.82 | 39,130.55 | 90 | 2,187,682 | 50,624,000 | 100.0 |
| 1971 | 8,100 | 6,683.52 | 45,581.66 | 51 | 2,781,385 | 54,137,000 | 100.0 |
| 1981 | 8,100 | 6,982.33 | 45,329.33 | 32 | 2,839,638 | 56,557,000 | 100.0 |
| 1991 | 8,100 | 7,009.63 | 42,450.26 | 31 | 2,775,250 | 56,778,000 | 100.0 |
| 2001 | 8,100 | 7,021.20 | 39,325.47 | 33 | 2,546,804 | 56,996,000 | 99.8 |

Table 3. Panel Unit Root Tests, Pesaran's CADF Statistic

| Cities (N) | US | Spain | Italy |
|------------|----------------|-----------------|----------------|
| 50 | -0.488 (0.313) | -0.915 (0.180) | 4.995 (0.999) |
| 100 | 0.753 (0.774) | 0.050 (0.520) | 5.983 (0.999) |
| 200 | 1.618 (0.947) | -2.866 (0.002) | -1.097 (0.136) |
| 500 | 1.034 (0.849) | -12.132 (0.000) | 5.832 (0.999) |

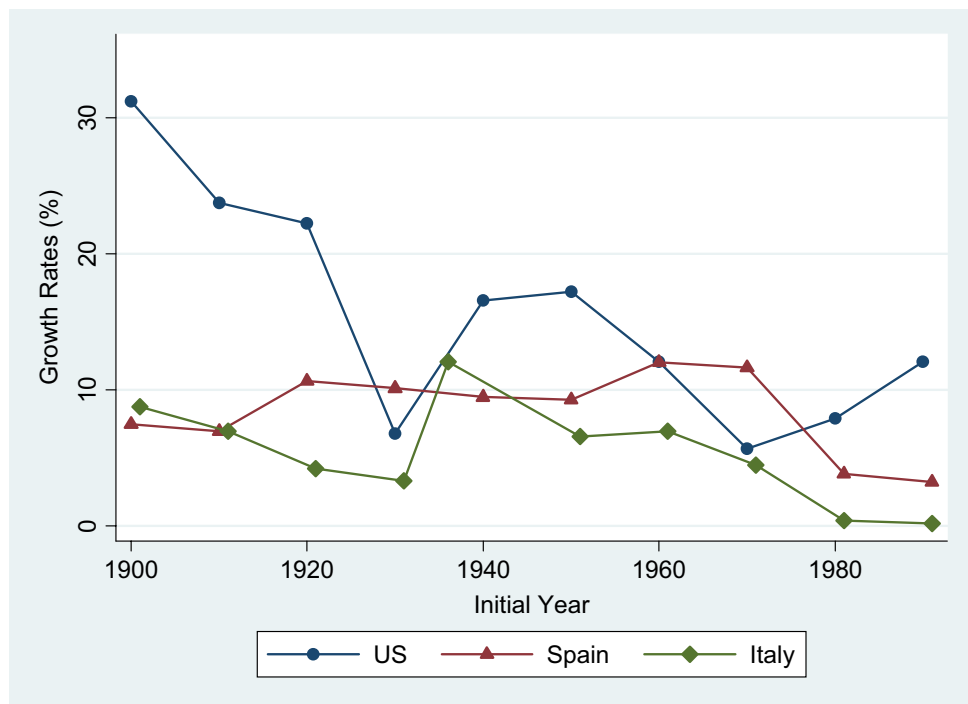
Notes: test-statistic (p-value). Pesaran's CADF test: standardised Ztbar statistic, $Z[\bar{t}]$. Variable: Relative size (in natural logarithms), sample size: (N, 11).

Table 4. Wilcoxon Rank-Sum Test of Lognormality by Year and Country

| US | | | | | | | | | | | |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Year | 1900 | 1910 | 1920 | 1930 | 1940 | 1950 | 1960 | 1970 | 1980 | 1990 | 2000 |
| p-value | 0.0252 | 0.017 | 0.0078 | 0.0088 | 0.0208 | 0.0464 | 0.1281 | 0.1836 | 0.2538 | 0.323 | 0.4168 |
| Spain | | | | | | | | | | | |
| Year | 1900 | 1910 | 1920 | 1930 | 1940 | 1950 | 1960 | 1970 | 1981 | 1991 | 2001 |
| p-value | 0.5953 | 0.6144 | 0.6233 | 0.6525 | 0.4909 | 0.5792 | 0.6049 | 0.522 | 0.5176 | 0.622 | 0.7212 |
| Italy | | | | | | | | | | | |
| Year | 1901 | 1911 | 1921 | 1931 | 1936 | 1951 | 1961 | 1971 | 1981 | 1991 | 2001 |
| p-value | 0.2081 | 0.2205 | 0.2352 | 0.291 | 0.2864 | 0.3118 | 0.2589 | 0.272 | 0.382 | 0.4671 | 0.5287 |

Ho: The distribution of cities follows a lognormal

Figure 1. Decennial Average Growth Rates by Country



Note: Average growth rates weighted by city size.

Figure 2. Scatter Plots of City Growth against City Size

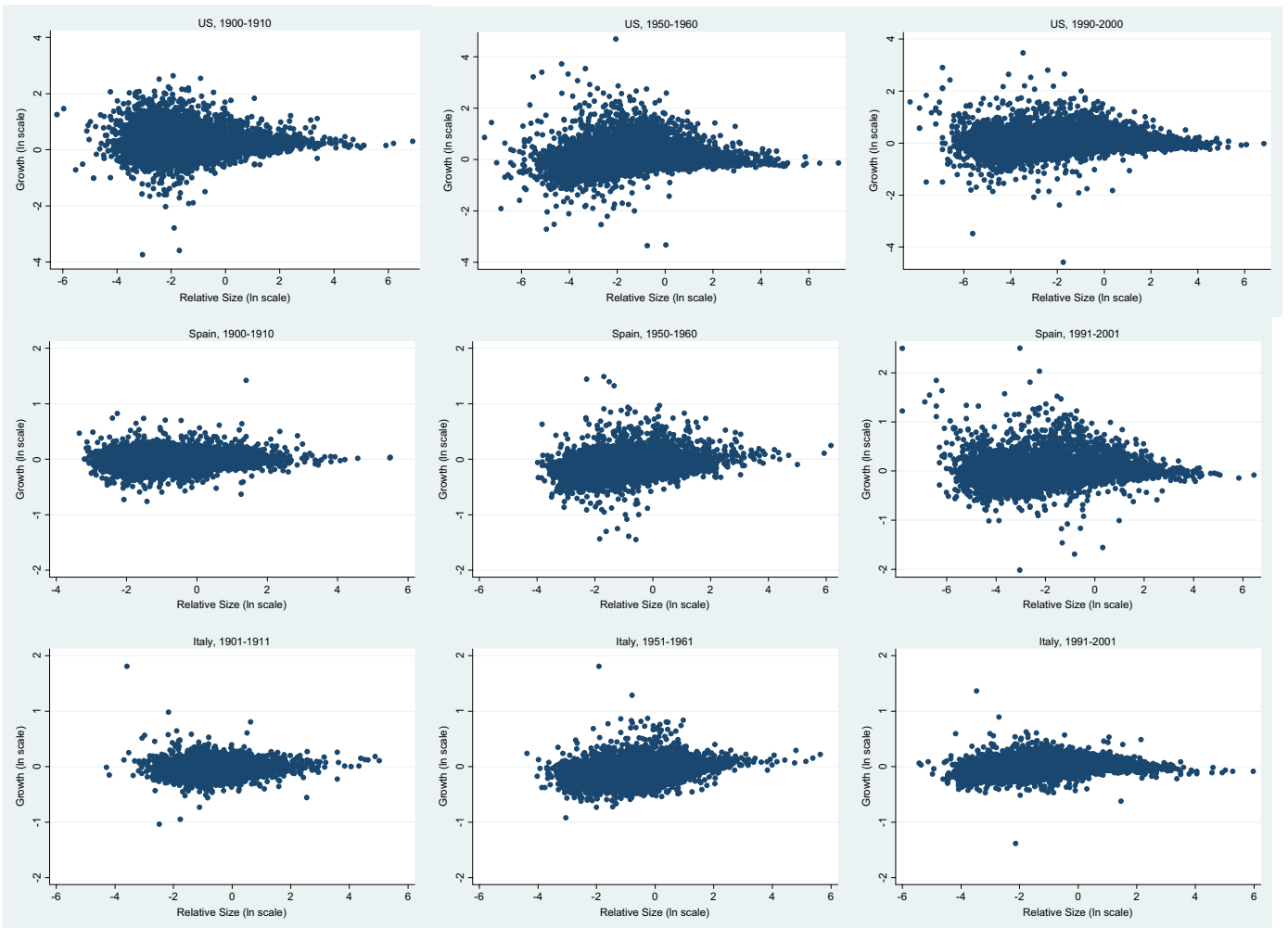


Figure 3. Nonparametric Estimates (Bandwidth 0.5) of the Growth Rate and its Variance for the US by Decade

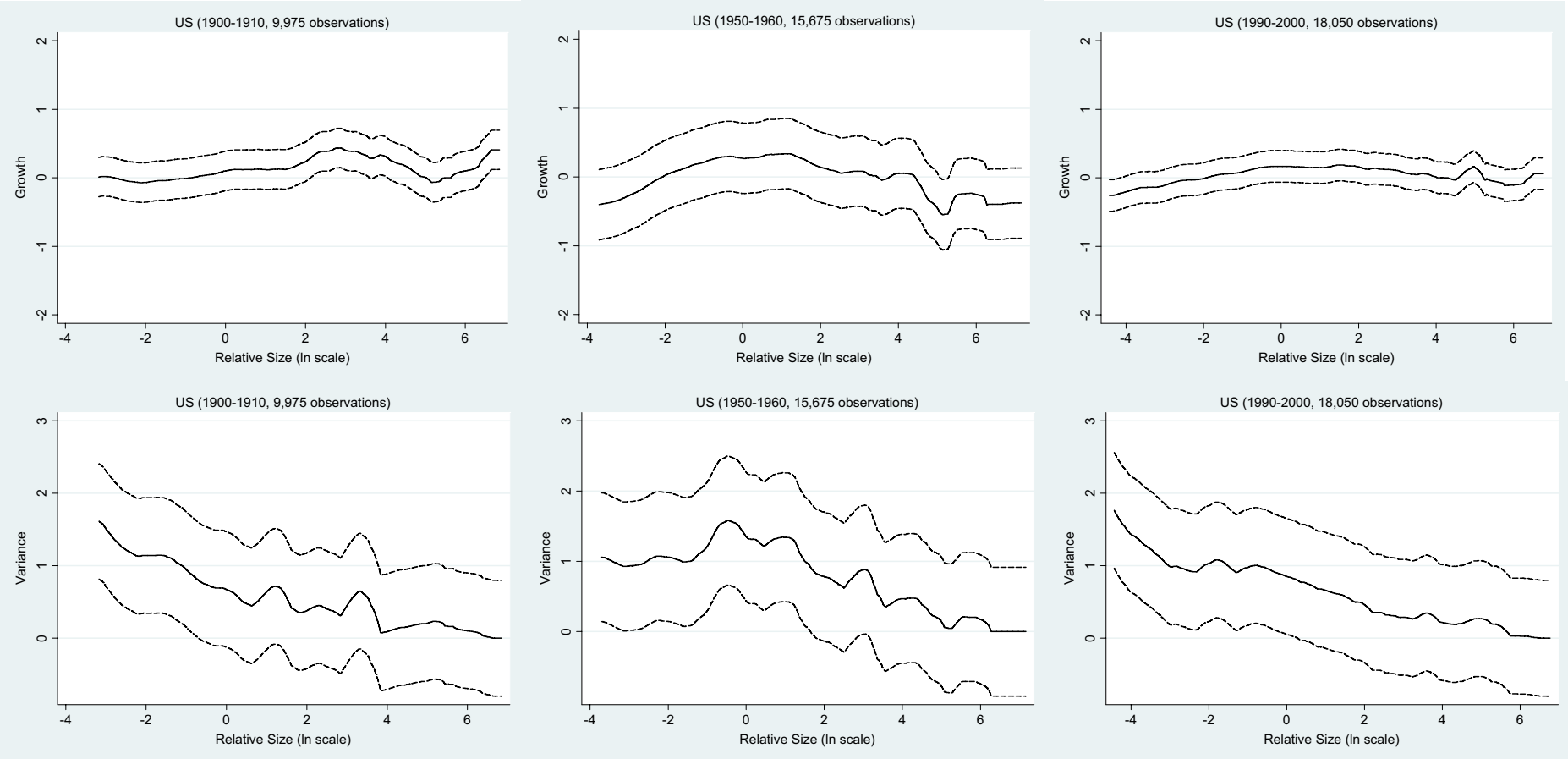


Figure 4. Nonparametric Estimates (Bandwidth 0.5) of the Growth Rate and its Variance for Spain by Decade

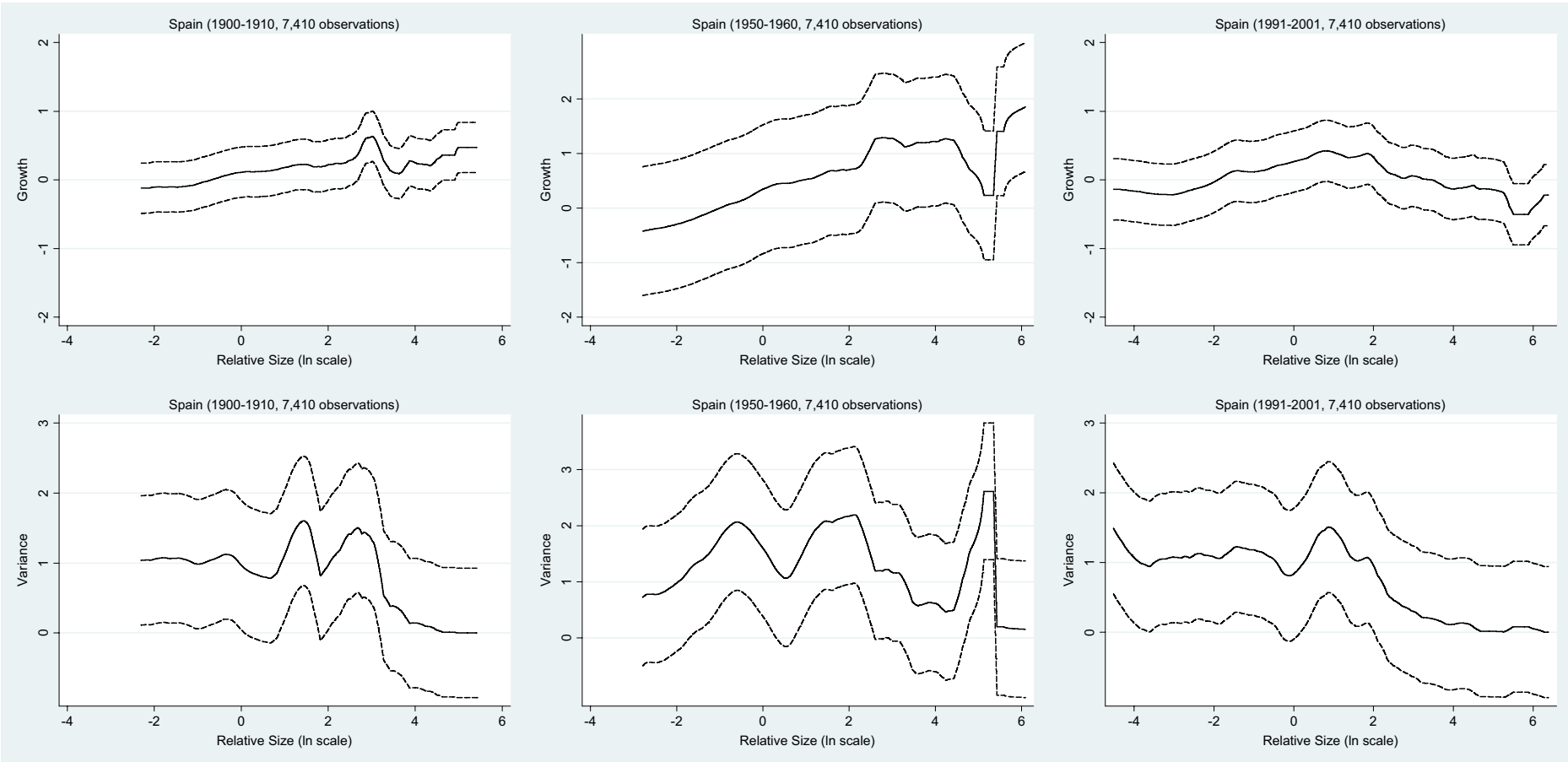


Figure 5. Nonparametric Estimates (Bandwidth 0.5) of the Growth Rate and its Variance for Italy by Decade

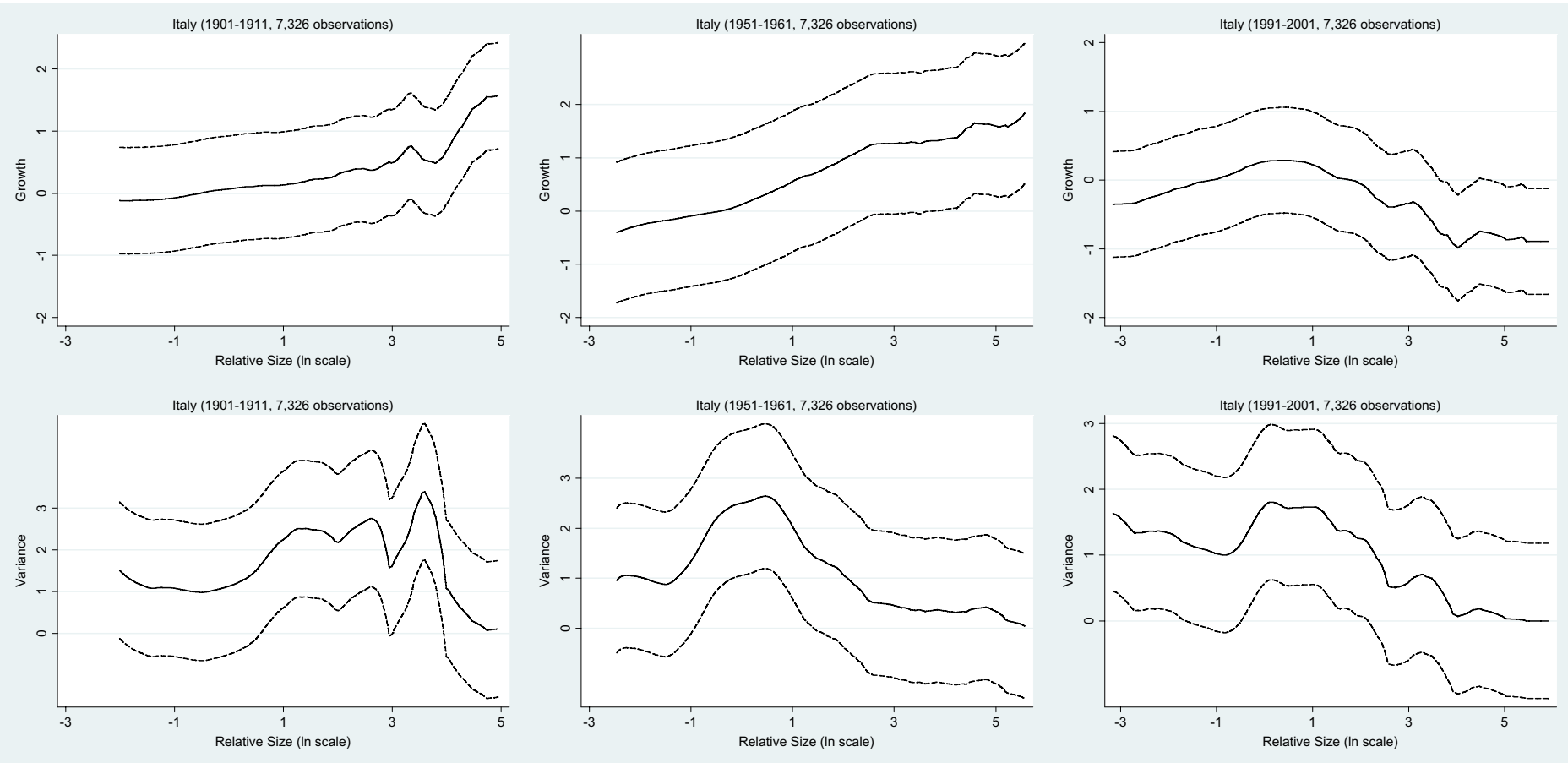


Figure 6. Nonparametric Estimates (Bandwidth 0.5) of the Growth Rate and its Variance. All the Twentieth Century

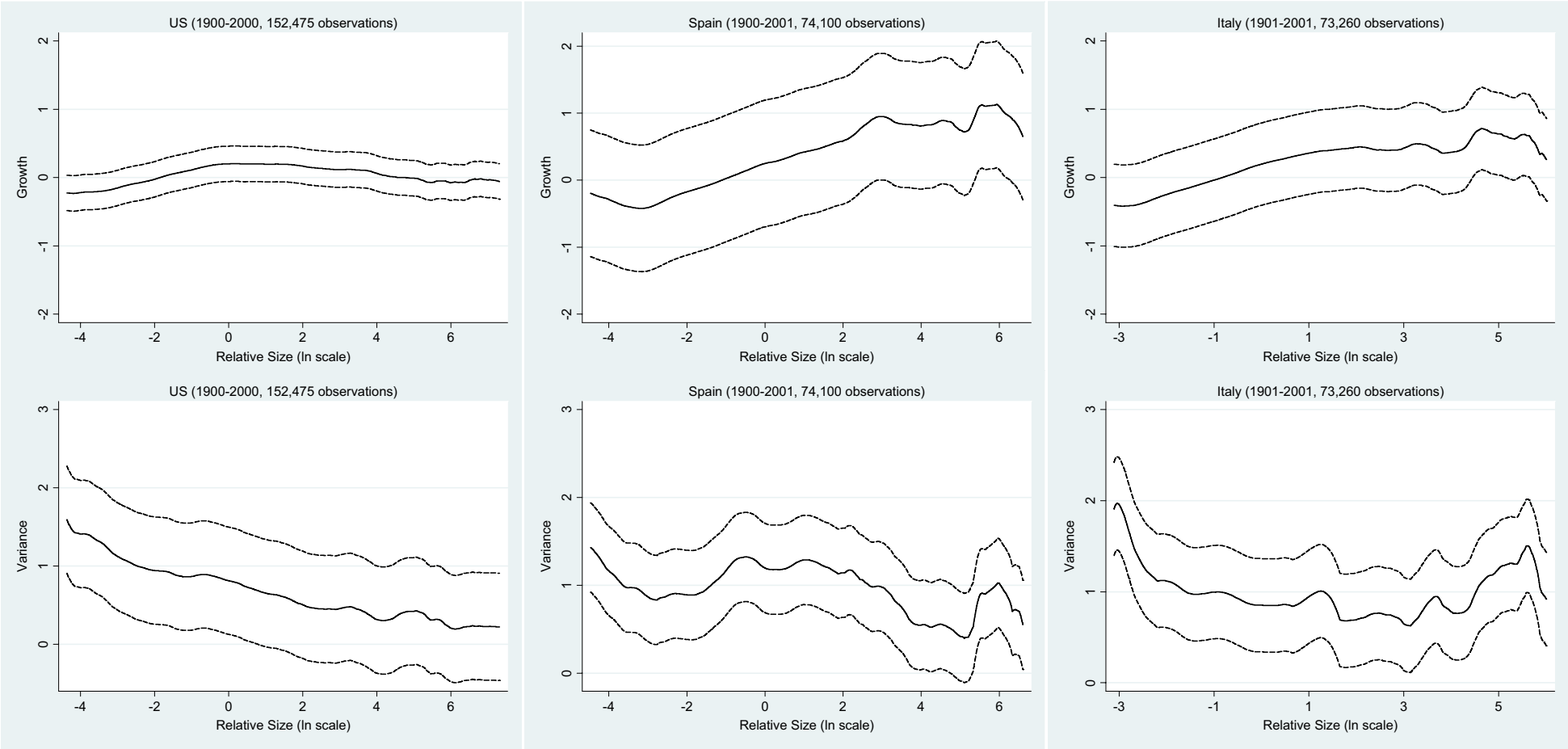


Figure 7. Nonparametric Estimates of the Growth Rate and its Variance. All the Twentieth Century, LOWESS (100% of the sample)

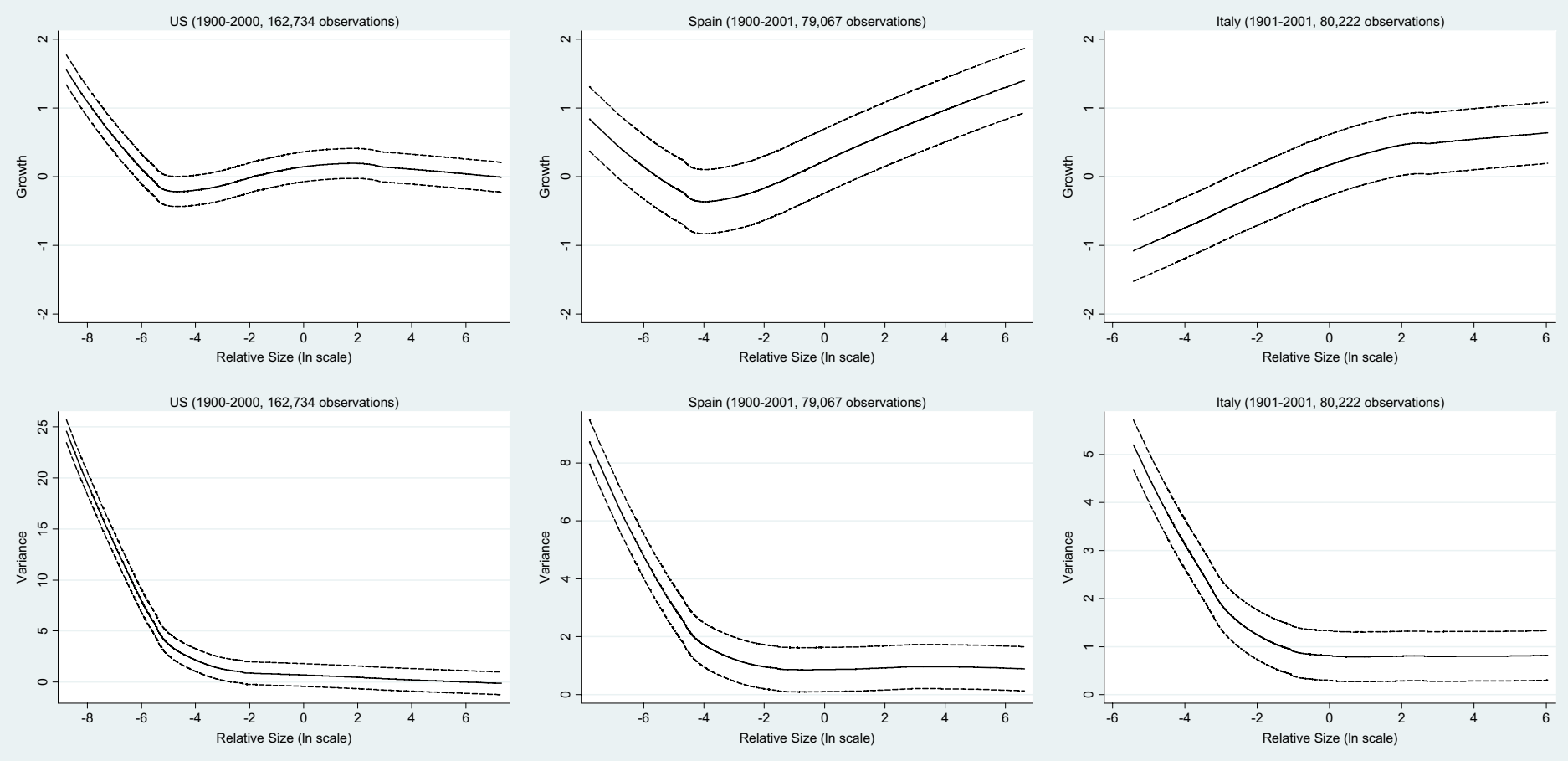
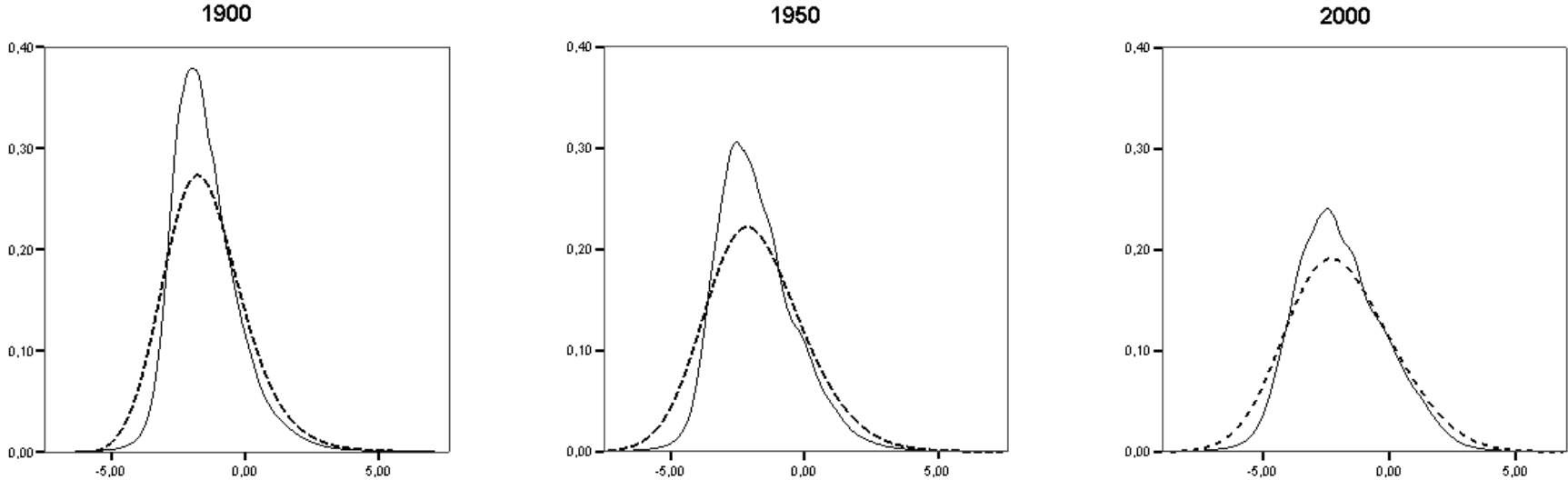


Figure 8. Estimated Density Function (In scale) and the Theoretical Lognormal (dotted line) for the US in 1900, 1950 and 2000



2010

- 2010/1, **De Borger, B., Pauwels, W.:** "A Nash bargaining solution to models of tax and investment competition: tolls and investment in serial transport corridors"
- 2010/2, **Chirinko, R.; Wilson, D.:** "Can Lower Tax Rates Be Bought? Business Rent-Seeking And Tax Competition Among U.S. States"
- 2010/3, **Esteller-Moré, A.; Rizzo, L.:** "Politics or mobility? Evidence from us excise taxation"
- 2010/4, **Roehrs, S.; Stadelmann, D.:** "Mobility and local income redistribution"
- 2010/5, **Fernández Llera, R.; García Valiñas, M.A.:** "Efficiency and elusion: both sides of public enterprises in Spain"
- 2010/6, **González Alegre, J.:** "Fiscal decentralization and intergovernmental grants: the European regional policy and Spanish autonomous regions"
- 2010/7, **Jametti, M.; Joanis, M.:** "Determinants of fiscal decentralization: political economy aspects"
- 2010/8, **Esteller-Moré, A.; Galmarini, U.; Rizzo, L.:** "Should tax bases overlap in a federation with lobbying?"
- 2010/9, **Cubel, M.:** "Fiscal equalization and political conflict"
- 2010/10, **Di Paolo, A.; Raymond, J.L.; Calero, J.:** "Exploring educational mobility in Europe"
- 2010/11, **Aidt, T.S.; Dutta, J.:** "Fiscal federalism and electoral accountability"
- 2010/12, **Arqué Castells, P.:** "Venture capital and innovation at the firm level"
- 2010/13, **García-Quevedo, J.; Mas-Verdú, F.; Polo-Otero, J.:** "Which firms want PhDs? The effect of the university-industry relationship on the PhD labour market"
- 2010/14, **Calabrese, S.; Epple, D.:** "On the political economy of tax limits"
- 2010/15, **Jofre-Monseny, J.:** "Is agglomeration taxable?"
- 2010/16, **Dragu, T.; Rodden, J.:** "Representation and regional redistribution in federations"
- 2010/17, **Borck, R.; Wimbersky, M.:** "Political economics of higher education finance"
- 2010/18, **Dohse, D.; Walter, S.G.:** "The role of entrepreneurship education and regional context in forming entrepreneurial intentions"
- 2010/19, **Åslund, O.; Edin, P-A.; Fredriksson, P.; Grönqvist, H.:** "Peers, neighborhoods and immigrant student achievement - Evidence from a placement policy"
- 2010/20, **Pelegrín, A.; Bolance, C.:** "International industry migration and firm characteristics: some evidence from the analysis of firm data"
- 2010/21, **Koh, H.; Riedel, N.:** "Do governments tax agglomeration rents?"
- 2010/22, **Curto-Grau, M.; Herranz-Loncán, A.; Solé-Ollé, A.:** "The political economy of infrastructure construction: The Spanish "Parliamentary Roads" (1880-1914)"
- 2010/23, **Bosch, N.; Espasa, M.; Mora, T.:** "Citizens' control and the efficiency of local public services"
- 2010/24, **Ahamdanech-Zarco, I.; García-Pérez, C.; Simón, H.:** "Wage inequality in Spain: A regional perspective"
- 2010/25, **Folke, O.:** "Shades of brown and green: Party effects in proportional election systems"
- 2010/26, **Falck, O.; Heblich, H.; Lameli, A.; Südekum, J.:** "Dialects, cultural identity and economic exchange"
- 2010/27, **Baum-Snow, N.; Pavan, R.:** "Understanding the city size wage gap"
- 2010/28, **Molloy, R.; Shan, H.:** "The effect of gasoline prices on household location"
- 2010/29, **Koethenbuerger, M.:** "How do local governments decide on public policy in fiscal federalism? Tax vs. expenditure optimization"
- 2010/30, **Abel, J.; Dey, I.; Gabe, T.:** "Productivity and the density of human capital"
- 2010/31, **Gerritse, M.:** "Policy competition and agglomeration: a local government view"
- 2010/32, **Hilber, C.; Lyytikäinen, T.; Vermeulen, W.:** "Capitalization of central government grants into local house prices: panel data evidence from England"
- 2010/33, **Hilber, C.; Robert-Nicoud, F.:** "On the origins of land use regulations: theory and evidence from us metro areas"
- 2010/34, **Picard, P.; Tabuchi, T.:** "City with forward and backward linkages"
- 2010/35, **Bodenhorn, H.; Cuberes, D.:** "Financial development and city growth: evidence from Northeastern American cities, 1790-1870"
- 2010/36, **Vulovic, V.:** "The effect of sub-national borrowing control on fiscal sustainability: how to regulate?"
- 2010/37, **Flamand, S.:** "Interregional transfers, group loyalty and the decentralization of redistribution"
- 2010/38, **Ahlfeldt, G.; Feddersen, A.:** "From periphery to core: economic adjustments to high speed rail"
- 2010/39, **González-Val, R.; Pueyo, F.:** "First nature vs. second nature causes: industry location and growth in the presence of an open-access renewable resource"
- 2010/40, **Billings, S.; Johnson, E.:** "A nonparametric test for industrial specialization"
- 2010/41, **Lee, S.; Li, Q.:** "Uneven landscapes and the city size distribution"
- 2010/42, **Ploeckl, F.:** "Borders, market access and urban growth; the case of Saxon towns and the Zollverein"
- 2010/43, **Hortas-Rico, M.:** "Urban sprawl and municipal budgets in Spain: a dynamic panel data analysis"
- 2010/44, **Koethenbuerger, M.:** "Electoral rules and incentive effects of fiscal transfers: evidence from Germany"

- 2010/45, Solé-Ollé, A.; Viladecans-Marsal, E.: "Lobbying, political competition, and local land supply: recent evidence from Spain"
- 2010/46, Larcinese, V.; Rizzo, L.; Testa, C.: "Why do small states receive more federal money? Us senate representation and the allocation of federal budget"
- 2010/47, Patacchini, E.; Zenou, Y.: "Neighborhood effects and parental involvement in the intergenerational transmission of education"
- 2010/48, Nedelkoska, L.: "Occupations at risk: explicit task content and job security"
- 2010/49, Jofre-Monseny, J.; Marín-López, R.; Viladecans-Marsal, E.: "The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms"
- 2010/50, Revelli, F.: "Tax mix corners and other kinks"
- 2010/51, Duch-Brown, N.; Parellada-Sabata M.; Polo-Otero, J.: "Economies of scale and scope of university research and technology transfer: a flexible multi-product approach"
- 2010/52, Duch-Brown, N.; Vilalta M.: "Can better governance increase university efficiency?"
- 2010/53, Cremer, H.; Goulão, C.: "Migration and social insurance"
- 2010/54, Mittermaier, F.; Rincke, J.: "Do countries compensate firms for international wage differentials?"
- 2010/55, Bogliacino, F.; Vivarelli, M.: "The job creation effect or R&D expenditures"
- 2010/56, Piacenza, M.; Turati, G.: "Does fiscal discipline towards sub-national governments affect citizens' well-being? Evidence on health"

2011

- 2011/1, Oppedisano, V.; Turati, G.: "What are the causes of educational inequalities and of their evolution over time in Europe? Evidence from PISA"
- 2011/2, Dahlberg, M.; Edmark, K.; Lundqvist, H.: "Ethnic diversity and preferences for redistribution "
- 2011/3, Canova, L.; Vaglio, A.: "Why do educated mothers matter? A model of parental help"
- 2011/4, Delgado, F.J.; Lago-Peñas, S.; Mayor, M.: "On the determinants of local tax rates: new evidence from Spain"
- 2011/5, Piolatto, A.; Schuett, F.: "A model of music piracy with popularity-dependent copying costs"
- 2011/6, Duch, N.; García-Estévez, J.; Parellada, M.: "Universities and regional economic growth in Spanish regions"
- 2011/7, Duch, N.; García-Estévez, J.: "Do universities affect firms' location decisions? Evidence from Spain"
- 2011/8, Dahlberg, M.; Mörk, E.: "Is there an election cycle in public employment? Separating time effects from election year effects"
- 2011/9, Costas-Pérez, E.; Solé-Ollé, A.; Sorribas-Navarro, P.: "Corruption scandals, press reporting, and accountability. Evidence from Spanish mayors"
- 2011/10, Choi, A.; Calero, J.; Escardíbul, J.O.: "Hell to touch the sky? private tutoring and academic achievement in Korea"
- 2011/11, Mira Godinho, M.; Cartaxo, R.: "University patenting, licensing and technology transfer: how organizational context and available resources determine performance"
- 2011/12, Duch-Brown, N.; García-Quevedo, J.; Montolio, D.: "The link between public support and private R&D effort: What is the optimal subsidy?"
- 2011/13, Breuillé, M.L.; Duran-Vigneron, P.; Samson, A.L.: "To assemble to resemble? A study of tax disparities among French municipalities"
- 2011/14, McCann, P.; Ortega-Argilés, R.: "Smart specialisation, regional growth and applications to EU cohesion policy"
- 2011/15, Montolio, D.; Trillas, F.: "Regulatory federalism and industrial policy in broadband telecommunications"
- 2011/16, Pelegrín, A.; Bolancé, C.: "Offshoring and company characteristics: some evidence from the analysis of Spanish firm data"
- 2011/17, Lin, C.: "Give me your wired and your highly skilled: measuring the impact of immigration policy on employers and shareholders"
- 2011/18, Bianchini, L.; Revelli, F.: "Green polities: urban environmental performance and government popularity"
- 2011/19, López Real, J.: "Family reunification or point-based immigration system? The case of the U.S. and Mexico"
- 2011/20, Bogliacino, F.; Piva, M.; Vivarelli, M.: "The impact of R&D on employment in Europe: a firm-level analysis"
- 2011/21, Tonello, M.: "Mechanisms of peer interactions between native and non-native students: rejection or integration?"
- 2011/22, García-Quevedo, J.; Mas-Verdú, F.; Montolio, D.: "What type of innovative firms acquire knowledge intensive services and from which suppliers?"

- 2011/23, Banal-Estañol, A.; Macho-Stadler, I.; Pérez-Castrillo, D.: "Research output from university-industry collaborative projects"
- 2011/24, Ligthart, J.E.; Van Oudheusden, P.: "In government we trust: the role of fiscal decentralization"
- 2011/25, Mongrain, S.; Wilson, J.D.: "Tax competition with heterogeneous capital mobility"
- 2011/26, Caruso, R.; Costa, J.; Ricciuti, R.: "The probability of military rule in Africa, 1970-2007"
- 2011/27, Solé-Ollé, A.; Viladecans-Marsal, E.: "Local spending and the housing boom"
- 2011/28, Simón, H.; Ramos, R.; Sanromá, E.: "Occupational mobility of immigrants in a low skilled economy. The Spanish case"
- 2011/29, Piolatto, A.; Trotin, G.: "Optimal tax enforcement under prospect theory"
- 2011/30, Montolio, D.; Piolatto, A.: "Financing public education when altruistic agents have retirement concerns"
- 2011/31, García-Quevedo, J.; Pellegrino, G.; Vivarelli, M.: "The determinants of YICs' R&D activity"
- 2011/32, Goodspeed, T.J.: "Corruption, accountability, and decentralization: theory and evidence from Mexico"
- 2011/33, Pedraja, F.; Cordero, J.M.: "Analysis of alternative proposals to reform the Spanish intergovernmental transfer system for municipalities"
- 2011/34, Jofre-Monseny, J.; Sorribas-Navarro, P.; Vázquez-Grenno, J.: "Welfare spending and ethnic heterogeneity: evidence from a massive immigration wave"
- 2011/35, Lyytikäinen, T.: "Tax competition among local governments: evidence from a property tax reform in Finland"
- 2011/36, Brühlhart, M.; Schmidheiny, K.: "Estimating the Rivalness of State-Level Inward FDI"
- 2011/37, García-Pérez, J.I.; Hidalgo-Hidalgo, M.; Robles-Zurita, J.A.: "Does grade retention affect achievement? Some evidence from Pisa"
- 2011/38, Boffa, f.; Panzar, J.: "Bottleneck co-ownership as a regulatory alternative"
- 2011/39, González-Val, R.; Olmo, J.: "Growth in a cross-section of cities: location, increasing returns or random growth?"
- 2011/40, Anesi, V.; De Donder, P.: "Voting under the threat of secession: accommodation vs. repression"
- 2011/41, Di Pietro, G.; Mora, T.: "The effect of the l'Aquila earthquake on labour market outcomes"
- 2011/42, Brueckner, J.K.; Neumark, D.: "Beaches, sunshine, and public-sector pay: theory and evidence on amenities and rent extraction by government workers"
- 2011/43, Cortés, D.: "Decentralization of government and contracting with the private sector"
- 2011/44, Turati, G.; Montolio, D.; Piacenza, M.: "Fiscal decentralisation, private school funding, and students' achievements. A tale from two Roman catholic countries"

2012

- 2012/1, Montolio, D.; Trujillo, E.: "What drives investment in telecommunications? The role of regulation, firms' internationalization and market knowledge"
- 2012/2, Giesen, K.; Suedekum, J.: "The size distribution across all "cities": a unifying approach"
- 2012/3, Foremny, D.; Riedel, N.: "Business taxes and the electoral cycle"
- 2012/4, García-Estévez, J.; Duch-Brown, N.: "Student graduation: to what extent does university expenditure matter?"
- 2012/5, Durán-Cabré, J.M.; Esteller-Moré, A.; Salvadori, L.: "Empirical evidence on horizontal competition in tax enforcement"
- 2012/6, Pickering, A.C.; Rockey, J.: "Ideology and the growth of US state government"
- 2012/7, Vergolini, L.; Zanini, N.: "How does aid matter? The effect of financial aid on university enrolment decisions"
- 2012/8, Backus, P.: "Gibrat's law and legacy for non-profit organisations: a non-parametric analysis"
- 2012/9, Jofre-Monseny, J.; Marín-López, R.; Viladecans-Marsal, E.: "What underlies localization and urbanization economies? Evidence from the location of new firms"
- 2012/10, Mantovani, A.; Vandekerckhove, J.: "The strategic interplay between bundling and merging in complementary markets"
- 2012/11, Garcia-López, M.A.: "Urban spatial structure, suburbanization and transportation in Barcelona"
- 2012/12, Revelli, F.: "Business taxation and economic performance in hierarchical government structures"
- 2012/13, Arqué-Castells, P.; Mohnen, P.: "Sunk costs, extensive R&D subsidies and permanent inducement effects"
- 2012/14, Boffa, F.; Piolatto, A.; Ponzetto, G.: "Centralization and accountability: theory and evidence from the Clean Air Act"
- 2012/15, Cheshire, P.C.; Hilber, C.A.L.; Kaplanis, I.: "Land use regulation and productivity – land matters: evidence from a UK supermarket chain"

2012/16, Choi, A.; Calero, J.: "The contribution of the disabled to the attainment of the Europe 2020 strategy headline targets"

2012/17, Silva, J.I.; Vázquez-Grenno, J.: "The ins and outs of unemployment in a two-tier labor market"



U
UNIVERSITAT DE BARCELONA
B

ieb@ub.edu
www.ieb.ub.edu

Cities and Innovation

Document de
treball de l'IEB