A century of shocks: The evolution of the German city size distribution 1925–1999

Maarten Bosker a,⁎, Steven Brakman b, Harry Garretsen a, Marc Schramm a

Utrecht University, Janskerkhof 12, 3512 BL, Utrecht, The Netherlands
University of Groningen, P. O. Box 800, 9700 AV, Groningen, The Netherlands

Article info
Article history:
Received 6 December 2006
Received in revised form 13 November 2007
Accepted 21 April 2008
Available online 17 May 2008

JEL classification:
O18
R0
N94

Keywords:
City size distributions
Urban growth
Zipf's Law
Gibrat's Law
Panel unit root tests

Abstract
This paper uses empirical evidence on the evolution and structure of the West-German city size distribution to assess the relevance of three different theories of urban growth. The West-German case is of particular interest as Germany's urban system has been subject to some of history's largest (exogenous) shocks during the 20th century. A unique annual data set for 62 West-German cities that covers the period 1925–1999 allows for the identification of these shocks and provides evidence on the effects of these ‘quasi-natural experiments’ on the city size distribution as a whole as well as on each city separately. Our main findings are twofold. First, WWII has had a major and lasting impact on the city size distribution. Second, and heavily based upon the results of (panel) unit root tests that analyze the evolution of the individual cities that make up the West-German city size distribution, city growth is found to be trend stationary, which is not in line with Gibrat’s Law of proportional effect. Overall, our findings are most consistent with theories emphasizing the role of increasing returns to scale for city growth.

© 2008 Elsevier B.V. All rights reserved.

JEL classification:
O18
R0
N94

Keywords:
City size distributions
Urban growth
Zipf's Law
Gibrat's Law
Panel unit root tests

1. Introduction

City size distributions and the underlying city size dynamics have received considerable attention in the urban economic literature in recent years. Empirical studies have in particular produced evidence with respect to three features of city size distributions. First, city size distributions are found to be remarkably stable over time. Second, the hierarchy of the individual cities making up these distributions is also often rather stable, which suggests proportionate city growth, see for example Eaton and Eckstein (1997) and Black and Henderson (2003). The third stylized fact is that city size distributions are very well approximated by a power law in the upper tail of the distribution. A special case of which is better known as Zipf’s Law and has been found to hold for various countries at various points in time, see e.g. Soo (2005) and Nitsch (2005).

These empirical regularities have stimulated the development of city growth models that can explain these features of city size distributions in a coherent economic framework. Modern theories (e.g. Eeckhout, 2004; Rossi-Hansberg and Wright, 2007; Córdoba, 2008) try to explain the evolution of city size distributions in a way that is consistent with the empirical results. Either a stable city size distribution adhering to Zipf’s Law in the upper tail follows directly from the model (Gabaix, 1999; Eeckhout, 2004), or it is one of the possible outcomes of the model (Rossi-Hansberg and Wright, 2007). These models have benefited substantially from the work of Gabaix (1999) who, building on earlier work by Simon (1955), showed that a stable city size distribution adhering

---

We would like to thank the editor, three anonymous referees, Stephen Redding, Daniel Sturm, and Joppe de Ree, and seminar participants at the London School of Economics, the 2006 NARSC meeting in Las Vegas, the 2006 EEA meeting in Vienna, Utrecht University, and Radboud University Nijmegen for very useful comments and suggestions and we also like to thank Peter Koudijs for excellent research assistance.

⁎ Corresponding author. Utrecht School of Economics, Utrecht University, Janskerkhof 12, 3512 BL Utrecht, The Netherlands.
E-mail address: m.bosker@econ.uu.nl (M. Bosker).

0166-0462/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
Recent empirical papers on the evolution of the city size distribution focus exclusively on the US experience, e.g. Black and Henderson (2003), Overman and Ioannides (2001), Ioannides and Overman (2003, 2004), Dobbins and Ioannides (2000, 2001), or Eeckhout (2004). Besides some simple Zipf studies that do not look at distributional dynamics nor at evidence for Gibrat’s Law of proportional effect, the only two papers we know of that offer a thorough look at the distributional dynamics of city size distributions for other countries than the USA are Eaton and Eckstein (1997) and Anderson and Ge (2005). The former provides evidence for France and Japan confirming the notion of a stable city size distribution, while the latter shows that in case of China the city size distribution has been affected in a predictable way by government policies.

The contributions of the present paper are the following. First, we examine the evolution of the city size distribution for West-Germany (from now on referred to as Germany). The interest in the German city size distribution can be dated back to as early as 1913 when the geographer Auerbach (1913), as one of the first, noted that the city size distribution could be approximated by a power law. For our empirical analysis we have constructed a unique data set of annual city population data for 62 of the largest cities in Germany over the period 1925–1999. This data set allows us to describe the evolution of the German city size distribution quite accurately. Second, the use of German data provides a specific empirical view on the evolution of the city size distribution, namely it offers a (unique) look at the effect of large shocks to the urban system. In the time period under consideration German cities were subject to a number of large ‘quasi-natural experiments’ namely the heavy destruction of cities during WW II and the split and subsequent reunification of Germany. Our data set allows us to look at the impact of these ‘quasi-natural experiments’ in a much more dynamic fashion than previous research on the same topic (Davis and Weinstein, 2002; Bosker et al., 2007; Redding and Sturm, 2005; Redding, Sturm, and Wolf, 2007). Third, our annual data set allows us to perform unit root tests for each individual city in order to find evidence on Gibrat’s Law of proportional effect. The present paper is among the first to provide (panel) unit root tests for cities that, given our extensive data set, arguably suffer less from low power problems that beset unit root tests based on a limited number of observations, while at the same time adequately controlling for the short-run dynamics in city size.

The results of our analysis of the German urban system can also help to distinguish between some of the proposed urban economic theories that try to explain the shape and evolution of city size distributions. Following Davis and Weinstein (2002), these theories can be grouped into three broad categories, i.e. increasing returns to scale, random growth and locational fundamentals. The models in all three categories predict a stable city size distribution in equilibrium; the reaction to shocks is however quite different. Models exhibiting increasing returns to scale can give rise to a stable distribution which is sensitive to shocks and which does not necessarily adhere to Zipf’s Law (see also Chapter 12 of Fujita et al., 1999; Gabaix and Ioannides, 2004; Brakman et al., 1999). As a result a large shock has the potential to (radically) change the city size distribution. Models falling under the random growth category, e.g. Gabaix (1999) or Córdoba (2008), predict that shocks have a permanent effect on city sizes, but given that these shocks are distributed randomly over cities and mean- and variance independent of city size, they will in the limit result in a city size distribution that adheres to Zipf’s Law in the upper tail. The effect of a large shock thus has no effect on the limiting city size distribution; it can however have a permanent impact on the relative position of cities within the distribution. Finally, the locational fundamentals approach suggests that the observed city size distribution is the result of fixed underlying locational fundamentals (first nature geography). A large shock will now result in both the city size distribution as a whole and the relative position of cities within this distribution returning to their pre-shock state. Given the three categories’ different reaction to large shocks, the ‘quasi-natural experiments’ that the German urban system was subjected to, provide a way to try and distinguish between competing views of city size evolution.

Our first main finding is that the German city size distribution is permanently affected by the World War II shock, more so than by any other shocks. Cities that have been hit relatively hard due to the substantial bombings and the subsequent allied invasion do not recover the loss in relative size. After the war, the city size distribution does not revert to its pre-WWII level, but shifts to one characterized by a more even distribution of population over the cities in the sample. Compared to the impact of WWII, the separation from and later reunion with East-Germany has had a much less severe impact on relative city sizes. Our second finding is that, once corrected for the heavy destruction during WWII, (panel) unit root tests that are used to test for the validity of
proportionate city growth reject Gibrat’s Law for about 75% of all cities. Also we find strong evidence that the locational fundamentals approach does not seem to explain the evolution of Germany’s urban system, which is in contrast to the findings of Davis and Weinstein (2002, 2004) for Japan. Overall the evidence does seem to comply best with urban theories exhibiting increasing returns to scale.

2. Data

In constructing our data set, we first had to choose which cities to include in our sample. We choose to include those West-German cities in our data set that either had a population of over 50,000 inhabitants before the beginning of WWII or cities that were over the sample period classified as Großstädte, cities with a population of at least 100,000 people. Cities are defined on a city-proper or administrative basis in Germany. Adjustments to the administrative boundaries (and hence size) of some cities were made during our sample period due to metropolitan developments but we were able to correct for the most significant of these changes (see Appendix A for more information on the city selection criteria and the boundary adjustments made). This selection process initially resulted in 81 cities, the same sample of West-German cities as used in Brakman et al. (2004) and Bosker et al. (2007). In those two studies the only requirement to be included in the analysis was that for each city, population data had to be available for the years 1939, 1946 and 1963 (and possibly 1933). In the present study, however, we only include cities for which we have annual population data for each year in the 1925–1999 period. In total 16 of the 81 West-German cities in our sample did not meet this requirement and for 3 cities we were unable to adequately correct for changes in their administrative border. We are therefore left with a data set that consists of 62 West-German cities over the period (Appendix A lists the cities included in our sample). The 19 cities that were dropped from our analysis were mostly relatively small. Since our main point of interest will be the upper tail of the city size distribution it can be argued that their exclusion does not matter too much for our analysis.

As to the decision to focus on West-German cities only, there are two main reasons to exclude East-German cities, that is to say cities that were part of the German Democratic Republic (GDR). The first reason is simply data availability. For most of the cities concerned there are too many missing observations during the GDR-period. The second and more fundamental reason is that (see Brakman et al., 2004) cities in the GDR were not part of the kind of urban system that lies at the heart of all urban location theories where economic agents are free to choose their location. On the contrary, in the centrally planned economy of the GDR, firms and workers were not free to move between cities. In our view this has the implication that any testing of the stability or any other feature of the pan-German city size distribution is not very useful during our sample period. Obviously, this does not imply that we are not concerned with the split between West and East Germany or the subsequent reunification, but we will deal with this from the perspective of West-German cities only. Finally, with respect to the length of the sample period one could argue that it may be worthwhile to include population data for the pre-1925 period as well so as to be able to deal with for instance the WWI shock. For some cities in our sample we have population data that go as far back as 1871, but the number of cities with annual pre-1925 data is rather small so we decided to take 1925 as our cut-off date.

3. Evolution of the city size distribution

We start our analysis by giving a description of the evolution of the West-German urban system. Table 1 below shows that during our sample period total population increased by about 70% from about 39 million people in 1925 to 67 million in 1999.

During the same period, the share of Germany’s population living in one of our sample cities declined by about 32% (or 13 ppt) suggesting a process of suburbanization over the sample period. The average city size in our sample increased by 16% from 258,000 in 1925 to 300,000 inhabitants but it has been quite stable from 1955 onwards. Comparing the development of mean city size to that of the median city size, which increased by 42%, the impression comes to the fore that smaller cities grew faster than the larger cities in our sample, hinting at a transition towards a more equally spread population over the cities in our sample. The impact of WWII also clearly shows up from Table 1. Whereas total West-German population increased by about 11%
between 1935 and 1945, the average population of the cities in our sample decreased by over 20%, indicating that the urban population in particular suffered substantial losses during WWII.

3.1. Distribution characteristics

Before going into the analysis of the evolution of the German city size distribution and in order to fix ideas, Fig. 1 shows the distribution for both the beginning (=1925) and the end (=1999) of our sample period. Also included in the figure are the city size distributions right at the start (=1939) and at the end (=1945) of WWII. In order to control for the changes in mean city size (see Table 1), we normalized city sizes for each year by dividing each city's population by the mean city size in the same year.

These kernel estimates of the city size distribution already reveal several interesting facts about the evolution of the German urban system. The first observation is that in the pre-WWII period from 1925–1939 the distribution remained remarkably stable; the two distributions overlap almost exactly. The second observation concerns the impact of WWII. The massive loss in urban population suffered during the war (see Table 1) induced a clear shift in the German city size distribution in a period of only 6 years. Comparing the 1945 distribution with the 1939 distribution, one can see that the main impact of the war is that the city size distribution loses mass in the lower tail and gains substantial mass in the middle. Keeping in mind that we are looking at the distribution of normalized city sizes this movement indicates that during the war the largest cities’ population grew slower (or maybe more appropriate, considering the war destruction, declined more) than that of the smaller cities in our sample. Finally, the last interesting point to make about these kernel estimations is that in the period after the war the city size distribution does not revert to its pre-war state; instead the movement that was initiated during the war seems to propagate itself with the distribution gaining even more mass in the middle and losing mass in the lower tail.

As a first glance at the dynamics within the distribution over our sample period, Fig. 2 plots cities’ initial against cities’ final rank within the city size distribution for the pre-WWII, the WWII, the post-WWII and the total sample period. Table 2 complements this by providing the cities that made the largest rank movement up and the largest rank movement down within the city size distribution respectively, as well as the absolute rank movement of the average city and the standard deviation of the average city’s (absolute) rank movement within the city size distribution.

Fig. 2 and Table 2 clearly show that cities’ intradistributional rank movements were considerably lower before WWII than during or after the war, conveying the same message of a major impact of the war on the urban landscape in Germany. Also the magnitude of the change in rank of the winner and the loser increases considerably during and after WWII. Over the whole sample period the average city moves four places up or down in the city size distribution. Interestingly, rank changes are more frequent (and larger) for smaller cities, which reflects the fact that the population differences between smaller cities are generally smaller so that a given population in- or decrease will more easily result in a rank change. Taken together, Table 2 and Figs. 1 and 2 suggest that both the movement of the distribution as a whole as well as the relative position of cities within the distribution are of importance.

---

5 The distributions are obtained by kernel estimation methods using a Gaussian kernel with the optimal bandwidth chosen using the method proposed in Silverman (1986).

6 Plotting the distributions for the years between 1925 and 1939 confirms this, but these are left out of the figure for the sake of clarity.

7 Plotting the distributions for the years between 1945 and 1999 confirms this pattern but are left out of the figure for the sake of clarity.
3.2. Distributional dynamics

To take a closer look at the (intra-)distributional dynamics suggested by Fig. 1 and Table 2, we now turn to the estimation of the movement of the city size distribution over the sample period. In order to do so, we use Markov chain techniques following Dobkins and Ioannides (2000), Black and Henderson (2003) and Eaton and Eckstein (1997). These techniques quantify the dynamics of the city size distribution as a whole based on the intradistributional dynamics of the individual cities that make up this distribution. The use of Markov chain techniques requires the quantification of the distribution by discretizing it, i.e. each city is assigned to one of a predetermined number of groups based on its relative size. Letting \( f_t \) denote the vector of the resulting discretized distribution at period \( t \) and assuming that the distribution follows a homogenous, stationary, first order Markov process, the distributional dynamics can be characterized by the following Markov chain,

\[
f_{t+x} = M f_t
\]

where \( M \) is the so-called \( x \)-period transition matrix that maps the distribution at period \( t \) into period \( t+x \). Each element \( m_{ij} \) in the transition matrix represents the probability that a city makes a move within the discretized distribution from group \( i \) in period \( t \) to group \( j \) in period \( t+x \). To discretize the city size distribution, we allocate each city to one of five groups based on its relative size. This requires the definition of the cut-off points that determine which city belongs to which group. Following Eaton and Eckstein (1997) and Quah (1993), we choose cut-off points exogenously and at city sizes of 0.25, 0.5, and times the average city size, \( \mu_t \), for a given year \( t \). Table 3 shows the resulting discretized distributions for the same years as for which Fig. 1 shows the kernel estimates of the entire empirical city size distribution.

Even though the distributions are substantially simplified by the discretization, the afore-mentioned pattern of stability before WWII and a shift towards the middle of the distribution during the war shows up in Table 3. Having discretized the distribution, we can now turn to the estimation of the transition matrix, \( M \). As we have yearly population data we choose to estimate the 1-year (\( x=1 \)) transition matrix. Each transition probability, \( m_{ij} \), in the transition matrix \( M \) is estimated by maximum likelihood along with its standard error, \( \sigma_{m_{ij}} \), i.e.

\[
\hat{m}_{ij} = \frac{\sum_{t=1}^{T-1} n_{i,j,t+1}}{\sum_{j=1}^{5} n_{i,j,t}} \quad \text{with} \quad \hat{\sigma}_{m_{ij}} = \sqrt{\frac{\hat{m}_{ij}(1-\hat{m}_{ij})}{N_i}}
\]

where \( n_{i,j,t+1} \) denotes the number of cities moving from group \( i \) in year \( t \) to group \( j \) in year \( t+1 \), \( n_{i,t} \) the number of cities in group \( i \) in year \( t \), and \( N_i = \sum_{j=1}^{5} n_{i,j,t} \). Using Eq. (2) we estimate the 1-year transition matrix for the pre-WWII period, the 6-year transition matrix during WWII (1939–1945) and the 1-year transition matrix for the post-WWII period.

Tables 4–6 show the corresponding estimates of these matrices. The diagonal elements of the 1-year transition matrices before (Table 4) and after (Table 6) the war are close to one, which indicates that the city size distribution does not change dramatically over a

---

\( ^8 \) Although quantitatively the results are sensitive to the choice of cut-off points this has no effect on the qualitative outcomes of our analysis.
### Table 2
Winners, losers and average absolute city rank movement

<table>
<thead>
<tr>
<th></th>
<th>Pre-WWII</th>
<th>WWII</th>
<th>Post-WWII</th>
<th>1925–1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max (−)</td>
<td>−5 (Pforzheim)</td>
<td>−17 (Würzburg)</td>
<td>−21 (Flensburg)</td>
<td>−15 (Wanne-Eickel)</td>
</tr>
<tr>
<td>Max (+)</td>
<td>5 (Stuttgart)</td>
<td>21 (Flensburg)</td>
<td>19 (Münster)</td>
<td>18 (Oldenburg)</td>
</tr>
<tr>
<td>Mean</td>
<td>1.7</td>
<td>4.7</td>
<td>3.7</td>
<td>4.0</td>
</tr>
<tr>
<td>SD</td>
<td>1.6</td>
<td>5.5</td>
<td>4.4</td>
<td>3.9</td>
</tr>
</tbody>
</table>

### Table 3
Discretized city size distributions

<table>
<thead>
<tr>
<th>City size [S]</th>
<th>1925</th>
<th>1939</th>
<th>1945</th>
<th>1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) S&lt;−0.25μ</td>
<td>0.129</td>
<td>0.129</td>
<td>0.065</td>
<td>0.016</td>
</tr>
<tr>
<td>2) 0.25μ&lt;−0.5μ</td>
<td>0.387</td>
<td>0.355</td>
<td>0.435</td>
<td>0.371</td>
</tr>
<tr>
<td>3) 0.5μ−0.5μ</td>
<td>0.210</td>
<td>0.242</td>
<td>0.226</td>
<td>0.355</td>
</tr>
<tr>
<td>4) μ&lt;−2μ</td>
<td>0.161</td>
<td>0.193</td>
<td>0.177</td>
<td>0.177</td>
</tr>
<tr>
<td>5) 2μ&lt;−S</td>
<td>0.113</td>
<td>0.081</td>
<td>0.097</td>
<td>0.081</td>
</tr>
</tbody>
</table>

Note: The numbers in the Table indicate the share of cities that fall in a particular category in a particular year. For example in 1925, 13% of the cities fell in the smallest category.

### Table 4
Pre-WWII 1-year transition matrix

<table>
<thead>
<tr>
<th>t/t+1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.983(0.012)</td>
<td>0.017(0.012)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.006(0.004)</td>
<td>0.988(0.006)</td>
<td>0.006(0.004)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0(0)</td>
<td>0.995(0.005)</td>
<td>0.005(0.005)</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.007(0.007)</td>
<td>0.986(0.009)</td>
<td>0.007(0.007)</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.034(0.019)</td>
<td>0.966(0.019)</td>
</tr>
</tbody>
</table>

Notes: 1, 2, …, 5 correspond to the different groups of the discretized distribution as in Table 3. The 2nd largest eigenvalue is 0.996. Standard errors between brackets.

### Table 5
WWII transition matrix

<table>
<thead>
<tr>
<th>1939/1945</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.250(0.153)</td>
<td>0.750(0.153)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.091(0.061)</td>
<td>0.818(0.082)</td>
<td>0.091(0.061)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.200(0.103)</td>
<td>0.733(0.114)</td>
<td>0.067(0.064)</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.083(0.080)</td>
<td>0.833(0.108)</td>
<td>0.083(0.080)</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1(0)</td>
</tr>
</tbody>
</table>

Notes: 1, 2, …, 5 correspond to the different groups of the discretized distribution as in Table 3. The 2nd largest eigenvalue is 0.992. Standard errors between brackets.

### Table 6
Post-WWII 1-year transition matrix

<table>
<thead>
<tr>
<th>t/t+1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.933(0.024)</td>
<td>0.067(0.024)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.003(0.001)</td>
<td>0.989(0.003)</td>
<td>0.008(0.002)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.004(0.002)</td>
<td>0.994(0.002)</td>
<td>0.002(0.001)</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.006(0.003)</td>
<td>0.982(0.006)</td>
<td>0.012(0.005)</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.018(0.007)</td>
<td>0.982(0.007)</td>
</tr>
</tbody>
</table>

Notes: 1, 2, …, 5 correspond to the different groups of the discretized distribution as in Table 3. The 2nd largest eigenvalue is 0.996. Standard errors between brackets.
period of 1 year. It is very interesting to note however that where before WWII all off-diagonal elements are not significantly different from zero, this changes after WWII when almost all off-diagonal elements are significantly different from zero. This significant off-diagonal movement indicates less stability of the distribution after the war, which complies with the visual inspection of Fig. 1. The magnitude and direction of this significant movement after the war can also be inferred when looking at the difference between the upper and lower off-diagonal entries of the transition matrix. This shows that most movement occurs from the smallest cities and, be it a lot smaller in magnitude, from the largest cities towards the middle of the distribution. This movement can be seen as evidence of a tendency of the distribution to gain mass in the middle indicating a city size distribution with more cities of medium size.

Next we turn to the impact of WWII on the German city size distribution. Table 1 already showed that the German urban population suffered a substantial loss during the war with a decrease of more than 10%. In turn the WWII-transition matrix in Table 5 shows the effect of this loss of urban population on the distribution in our sample. The most striking result is that 75% of the cities in the smallest category in 1939 made the transition to the 2nd to smallest category during the war period, indicating that the smallest cities suffered substantially less than the average city during the war. Second in magnitude is the finding that 20% of the cities in the middle category moved one category down to the 2nd to smallest group, indicating that small–medium sized cities suffered quite substantial losses during the war. Not a single city in the highest category shifts to a lower category due to the destruction during the war, this however does not indicate that these cities were not hit very hard during the war; it merely reflects the fact that these cities were very large before the war and the substantial loss of population during the war was not large enough to make them shift to a lower category in the discretized distribution.

Besides the fact that the transition matrices themselves are of interest, one can also use them to do interesting thought experiments (see initially Ioannides and Overman, 2004, but also Black and Henderson, 2003; Eaton and Eckstein, 1997). Using our estimated pre- and post-WWII-transition matrices we ask ourselves the following questions:

1. Assuming WWII would not have happened and the transition matrix remained as in the pre-WWII period, what would the city size distribution have looked like in 1945 and 1999?
2. Assuming that the estimated transition matrix remained as during the pre-WWII (post-WWII) period, would the city size distribution converge and if so what would it look like in the steady state?

To answer the first question, we use the estimated pre-WWII transition matrix from Table 4 and the observed distribution in 1925 from Table 3 to predict the distribution in 1945 and 1999, using the following formula,

$$\hat{f}_{1925+x} = M_{pre}^x f_{1925}$$

where $f_{1925+x}$ is the distribution $x$ years from the observed distribution in year 1925 and $M_{pre}^x$ is the pre-WWII transition matrix multiplied $x$ times by itself, e.g. $M_{pre}^2 = M_{pre} \times M_{pre}$. Column 1 and column 3 of Table 7 below give the resulting predicted distributions for 1945 and 1999 (taking $x=20$ or 74) respectively.

Comparing these two with the actual observed distributions for these years in Table 3, we see that for 1945 the main difference between the predicted—as if WWII did not happen—distribution and the actual distribution is found in the smallest two groups. The smallest group is predicted much too large, about 200% as large as observed and the second smallest group much too small, about 25% that of which observed. Comparing the actual and predicted—as if WWII did not happen—distribution for 1999, one sees a continuation of this pattern with the smallest group being predicted 10 times as large as observed and the second smallest group predicted at about two thirds the actual size. This confirms the notion that during WWII the smallest cities in the distribution grew faster, or better suffered a lower loss of population, than the average German city, and a continuation of this pattern after the war.

The same conclusion can be drawn from our second thought experiment: what would happen to the city size distribution if it continued to evolve as estimated by either the 1-year pre-WWII or 1-year post-WWII transition matrix? If one is willing to accept this continuation assumption, these limiting distributions are shown in columns 3 and 4 for the pre-WWII and the post-WWII case respectively. The pre-WWII limiting distribution is not that informative because it assigns zero mass to the two smallest groups, which is due to the fact that the pre-WWII transition matrix gives zero probability to a city in one of the three highest categories to make the transition to one of the two smallest categories (see Table 4). As cities in the two smallest groups do eventually transfer to

---

Table 7
Predicted city size distributions

<table>
<thead>
<tr>
<th>City size (S)</th>
<th>1945</th>
<th>1999</th>
<th>Limit</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition matrix</td>
<td>$M_{pre}$</td>
<td>$M_{pre}$</td>
<td>$M_{pre}$</td>
<td>$M_{post}$</td>
</tr>
<tr>
<td>Observed distribution</td>
<td>$F_{1925}$</td>
<td>$F_{1925}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>0.130</td>
<td>0.115</td>
<td>0</td>
<td>0.010</td>
</tr>
<tr>
<td>2</td>
<td>0.341</td>
<td>0.257</td>
<td>0</td>
<td>0.237</td>
</tr>
<tr>
<td>3</td>
<td>0.255</td>
<td>0.347</td>
<td>0.515</td>
<td>0.485</td>
</tr>
<tr>
<td>4</td>
<td>0.199</td>
<td>0.232</td>
<td>0.404</td>
<td>0.160</td>
</tr>
<tr>
<td>5</td>
<td>0.075</td>
<td>0.049</td>
<td>0.081</td>
<td>0.108</td>
</tr>
</tbody>
</table>

Notes: $M_{pre}$, $M_{post}$ denote the estimated 1-year transition matrix for the pre- and post-WWII period transition matrix (see Tables 3 and 5) respectively. 1, 2, ..., 5 correspond to the different groups of the discretized distribution as in Table 3.
the higher categories, this will in the limit result in an emptying of these two categories. It does however suggest a movement towards a city size distribution characterized by medium–large cities. The limiting distribution based on the transition matrix after the war gives a completely different picture, namely that of a city size distribution characterized by small–medium sized cities, i.e. a more equal spreading of population over the urban landscape.

The overall impression from the above analysis of both the intradistributional and distributional dynamics of the German city size distribution is that WWII did have a substantial direct and lasting effect on the German urban landscape. In the Introduction of our paper we stated that theories on urban growth could be distinguished according to their prediction regarding the stability of the city size distribution to shocks. Based on the evidence presented in this section our conclusion is that the (evolution of the) German (i.e. West-German) city size distribution during the period 1925–1999 has been sensitive to the WWII shock both in terms of the city size distribution as a whole as the relative position of cities within the distribution. So far our analysis has been largely descriptive however, in the remainder of our paper we will therefore turn from describing the evolution of the city size distribution to providing empirical tests that allow us to distinguish more properly between the three competing theories mentioned in the introduction.

4. Zipf’s Law and Gibrat’s Law of proportional effect

As already mentioned in the Introduction, the notion of a power law distribution describing the upper tail of city size distribution goes at least as far back as 1913 when the German geographer Auerbach (1913) noted this to be the case for Germany. The empirical literature has mainly focused on a special case of such a power law, namely that of city sizes in the upper tail of the distribution being distributed Pareto with coefficient \( \alpha = 1 \). This empirical regularity is better known as Zipf’s Law (see Zipf, 1949) and has been found to hold approximately for many countries over several years (see e.g. Soo, 2005 and Nitsch, 2005). The studies that test for Zipf’s Law mainly do so by means of a Zipf regression, that is regressing the log of cities’ rank on the log of city sizes. If city sizes are indeed distributed according to a power law it can be easily shown (see e.g. Eeckhout, 2004) that the rank of a particular city in this distribution is given by:

\[
r_i = N \left( \frac{S_i}{S_o} \right)^a
\]

where \( S_i \) is the size of city \( i \), \( S_o \) is a (arbitrary) minimum city size, and \( N \) the number of cities above this truncation point. Testing for Zipf’s Law can now be done by a regression of \( \ln(r) \) on \( \ln(S) \) in order to estimate the so-called Zipf coefficient, i.e. rewriting Eq. (4) in logs gives:

\[
\ln(r_i) = \alpha \ln S_i + \epsilon_i
\]

where \( \alpha = \ln(N) + \ln(S_o) \) is a constant and \( \epsilon_i \) a random error term. If it cannot be rejected that the estimated \( \hat{\alpha} \) equals 1 this constitutes evidence in favor of Zipf’s Law. We estimated Eq. (5) using OLS for all years in our sample separately:

\[
\ln(r_{it}) = 0.5 = \alpha_t - \alpha_t \ln(S_{it}) + \epsilon_{it}
\]

where we add \( -0.5 \) to all city ranks following the recommendation in Gabaix and Ibragimov (2007). This ensures that we get unbiased estimates of \( \alpha \). Fig. 3 below shows the results for \( \alpha_t \) and its corresponding 5% confidence interval (\( \hat{\alpha} \pm 2\hat{\sigma}_\alpha \)) corresponding to the estimated OLS standard error of \( \alpha_t \), \( \hat{\sigma}_\alpha \), for each of the years in our sample.

Fig. 3 shows that during the pre-WWII period the point estimate of \( \alpha \) is very close to 1. The impact of WWII also shows clearly when looking at the Zipf-regression results. The estimated Zipf coefficient, \( \hat{\alpha} \), increases from 1.09 in 1939 to about 1.16 right after the war, again confirming the notion of a more equal spread of urban population over the West-German cities in our sample due to the relative lower loss of urban population of smaller cities during WWII. In the post-WWII period the point estimate shows no return to 1, instead it steadily increases, finally reaching a point estimate of 1.20 in 1999.

These results are consistent with the idea that WWII has had a major impact, shocking the city size distribution from one adhering closely to Zipf’s Law to one characterized by a more equal spread of urban population over the different cities. This immediate impact of the war is not reversed in the post-WWII period, instead the distribution moves to one characterized by an even more equal spread of urban population over the West-German cities in our sample due to the relative lower loss of urban population of smaller cities during WWII. Based on the evidence presented in this section our conclusion is that the (evolution of the) German (i.e. West-German) city size distribution during the period 1925–1999 has been sensitive to the WWII shock both in terms of the city size distribution as a whole as the relative position of cities within the distribution. So far our analysis has been largely descriptive however, in the remainder of our paper we will therefore turn from describing the evolution of the city size distribution to providing empirical tests that allow us to distinguish more properly between the three competing theories mentioned in the introduction.

10 A likelihood ratio test for the time-homogeneity of the transition probabilities before and after WWII, also confirms this notion. (LR-statistic, distributed \( \chi^2[8] \): 70.70 [\( p \)-value 0.000]).

11 This is consistent with the findings in Brakman et al. (2004) and Bosker et al. (2007), see also Section 5 below.

12 Formally a variable (in our case city size), \( S \), adhering to a power law is distributed according to a Pareto distribution if the density function of this variable satisfies, \( p(S) = \frac{\alpha}{S^\alpha} \) for \( S \geq S_o \).

13 Approximately is the key word here, see also Brakman et al. (2001), ch. 7 or Gabaix and Ioannides (2004) for evidence that the Zipf coefficient is sometimes significantly different from 1 and/or changing over time.

14 As was pointed out by Gabaix and Ioannides (2004), there are some pitfalls when estimating a Zipf regression by OLS. The standard errors of the estimated \( \hat{\alpha} \) are also typically underestimated leading to an overrejection of Zipf’s Law. We keep to OLS noting that using the approximate standard errors as suggested by Dobkins and Ioannides (2000), however as noted in Embrechts et al. (1997) and Gabaix and Ioannides (2004) the small sample properties of this estimator are particularly bad and we therefore decided not to use it for our sample of 62 cities. Results not shown in the paper are all available upon request.
4.1. Individual city size evolution and Gibrat’s Law of proportional effect

Gabaix (1999) and also Eeckhout (2004) opened up an alternative way to empirically verify the relevance of Zipf’s Law for a particular city size distribution. Here we follow Gabaix (1999) who showed that if individual city size growth adheres to Gibrat’s Law of proportional effect, i.e. if city sizes grow randomly with the same expected growth rate and the same variance independent of city size, and is also subject to a lower reflective boundary (where smaller cities that hit the boundary from above are ‘bounced back upward’), the city size distribution converges to one adhering to Zipf’s Law. It is the combination of a random walk process with a lower reflective boundary that gives a Pareto distribution with a coefficient of one, i.e. Zipf’s Law. Without this lower reflective boundary, the city size distribution as a whole would in this case be log-normally distributed (see Eeckhout, 2004). As already briefly mentioned in the introduction, the fact that proportionate city growth gives rise to a lognormal city size distribution is also frequently referred to as Gibrat’s Law (see e.g. Sutton, 1997; Eeckhout, 2004). To avoid confusion, whenever we refer to Gibrat’s Law from now on, we mean, as most authors in the modern city size distribution literature (see e.g. Gabaix, 1999; Gabaix and Ioannides, 2004, or Black and Henderson, 2003), Gibrat’s Law of proportional effect. Gabaix’s (1999) contribution, showing that Gibrat’s Law combined with a lower reflective boundary leads to Zipf’s Law in (the upper tail of) the overall city size distribution, provides a link with the results we showed in the previous section. In the next sections we look for empirical evidence on Gibrat’s Law using both non-parametric and parametric techniques.

4.2. Non-parametric evidence on Gibrat’s Law

Ioannides and Overman (2003) and Eeckhout (2004) resort to non-parametric multivariate kernel estimations to shed empirical light on the relevance Gibrat’s Law. Both papers find considerable evidence in favor of Gibrat’s Law in case of the US urban system. Following this methodology, we plot the distribution of five-year city size growth rates conditional on initial city size (in logs) for the total, pre-WWII, post-WWII and WWII period along with the corresponding contour plots. The results are shown in Fig. 4a–d below. By taking any point, say S, on the city size axis and taking the cross-section through the kernel estimate parallel to the city size growth rate axis, one obtains the distribution of five-year growth rates conditional on the log of initial city size being S.

The plotted kernel in Fig. 4a shows that before WWII, German city size growth behaved remarkably well according to Gibrat’s Law: the estimated kernel and corresponding contour plot show no significant difference in the conditional growth rate distribution for cities of different size. This provides evidence that the growth rate of the cities in our sample in the pre-WWII period did not depend on the initial size of the city. It is interesting to relate these results to our estimated Zipf regression. The fact that before WWII individual city growth seems to adhere quite well to Gibrat’s Law corresponds nicely to the finding of a Zipf coefficient very close to 1 during that period (see Fig. 3).

After WWII, the evidence from the estimated stochastic kernel and contour plot in Fig. 4c shows a completely different picture. For the largest cities in our sample city growth rates still seem largely independent of initial city size. This does not hold true for the

---

15 For our present purposes we are not as such interested in the question whether the German city size distribution displays log-normality. Instead we want to find out if in the German case individual city growth is driven by a process of proportional growth. Also our dataset only covers the upper tail of the city size distribution, so that we are wary of making statements about the log-normality of the city size distribution as a whole. As pointed out by one of our referees, the normalized city size distributions shown in Fig. 1 already suggest that city-sizes are neither log-normally nor Pareto distributed. Indeed, when we test for this using various standard tests for log-normality (skewness kurtosis, Shapiro Wilk or Shapiro Francia) or the ’Paretoness’ of the distribution (Kolmogorov–Smirnov), we invariably reject the city size distribution to be log-normal or Pareto for the 4 years shown in Fig. 1 (results available upon request).

16 The stochastic kernels are estimated non-parametrically using a Gaussian kernel and with the bandwidth chosen following Silverman (1986). The contour plots can be read in the same way as standard topographical height maps with the lines in the plots connecting points on the distribution of similar height.
smaller cities in our sample however. Fig. 4c shows clear evidence that the smaller cities in our sample grow on average faster than the larger cities. This shift in the mean of the conditional distribution for smaller cities is most clearly seen from the contour plot, which makes a significant shift to the right for the smallest cities. The evidence for Gibrat’s Law is thus weaker for the post-WWII period, reflecting itself in the higher Zipf coefficient in Fig. 3.

Before turning to the overall picture of individual German city growth over the whole sample period from 1925–1999, we turn to the immediate impact of WWII on German city sizes. As can be seen in Fig. 4b, the largest cities suffered more heavily during the 6-year period of WWII. Smaller cities suffered less on average but the variance of city growth during the war also increases substantially for the smaller cities in the sample. This suggests that larger cities were not only damaged more heavily, the destruction of these larger cities was also less variable; or put differently, the larger cities were all hit quite similarly and more heavily compared to the smaller cities in our sample (evidence of the success of the Area Bombing-tactics of Allied Bomber Command aimed at destroying Germany’s urban centers, see Brakman et al., 2004).
Combining Fig. 4a–c, the estimated stochastic kernel for the whole sample period in Fig. 4d does not provide very clear evidence that Gibrat’s Law holds. The effect of the lesser destruction of smaller cities during WWII, and the higher growth rates of these same smaller cities after the war, dominates the equal growth rates before the war. As a result, the overall picture shows a similar (be it somewhat smaller) shift in the conditional growth rate distribution for the smaller cities in the sample as for the post-WWII case. This result of a size dependent mean growth rate, and a more or less initial size independent variance contrasts to empirical studies done using US city size data. Both Eeckhout (2004) and Ioannides and Overman (2003) find evidence for the USA of a more or less size independent mean growth rate and an increasing variance for smaller cities, which can be reconciled with the decreasing Zipf coefficient they find for the USA. The explanation of the higher Zipf coefficient found in our sample over the post-WWII period is a more equal spreading of urban population over the cities in Germany and not so much an increased variance of city growth for smaller cities.

5. Unit root testing and parametric evidence on Gibrat's Law

The non-parametric kernel estimates of the previous section remain largely based on pooled panel data evidence. In order to fully exploit the time series dimension of our data set, we now turn to a, more dynamic way of testing for Gibrat's Law. Given the many observations over time in our data set, we follow the suggestion made by Gabaix and Ioannides (2004) who state: "Hence
one can imagine that the next generation of city size evolution empirics could draw from the sophisticated econometric literature on unit roots” (Gabriel and Ioannides, 2004, pp. 20).

Clark and Stabler (1991) were one of the first to notice that Gibrat’s Law can be tested using unit root tests. Following their exposition of the relationship between Gibrat’s Law and unit root testing, assume that the size of city $i$ at time $t$, can be related to the size of that same city at time $t-1$ according to the following formula:

$$S_{it} = \gamma_i S_{it-1}$$

(7)

where $\gamma_i$ denotes the growth rate of city $i$ over the period $t-1$ to $t$. Next assume that this growth rate can be decomposed into three components, a random component $\epsilon_{it}$, a non-stochastic component relating the current growth rate to a (possibly time-varying) constant, past growth rates and initial city size:

$$\gamma_{it} = K_0 S_{it-1} \cdot \prod_{j=1}^{p-1} \gamma_{it-j} (1 + \epsilon_{it})$$

(8)

where $K_0$ is a possibly time-varying constant, and $\delta_i$ and $\beta_{ij}$ are parameters measuring the relative importance of initial city size and past growth rates on current city growth respectively and $\epsilon_{it}$ is a random error term. Now Gibrat’s Law would require $\delta_i=0$, such that initial city size does not influence the growth of a particular city. In order to be able to test for this, substitute Eq. (8) into Eq. (7), take logs and subtract $\ln S_{it-1}$ from both sides of the equation to obtain the following estimatable equation:

$$\Delta \ln S_{it} = c_i + \rho_i \Delta \ln S_{i,t-1} + \sum_{j=1}^{p} \beta_{ij} \Delta \ln S_{i,t-j} + \epsilon_{it}$$

(9)

where $c_i=K_0$ and the following approximate equality is used: $\ln(1 + \epsilon_{it}) \approx \epsilon_{it}$ for small values of $\epsilon_{it}$. This shows immediately that testing for Gibrat’s Law amounts to testing for a unit root in city sizes. If we find that $\rho_i$ is not significantly different from zero, i.e. a unit root in city size, this constitutes evidence in favor of city $i$‘s growth rate being independent of city $i$’s size. On the other hand an estimated $\rho_i$ smaller than zero would indicate that the evolution of city $i$ is a stationary process implying that city $i$’s growth rate declines with initial city size.

There are to date very few studies that actually perform unit root tests on individual city sizes. Notable exceptions are Clark and Stabler (1991), Black and Henderson (2003) and Sharma (2003). Clark and Stabler (1991) conclude in favor of the relevance of Gibrat’s Law based on the evolution of the city sizes of the seven largest Canadian cities over the period 1975–1984. As the ADF unit root tests are infamous for their small-sample properties, a sample period of only 10 years seems to put a substantial doubt on their results.17 Black and Henderson (2003) find no evidence for Gibrat’s Law using data on the US metropolitan areas when testing for a unit root in city sizes. As Clark and Stabler (1991) they also have only 10 observations over time (decade-by-decade city sizes over a total period of 90 years). They take explicit note of the hereby potentially induced small-sample problem and resort to a recently proposed panel unit root test, i.e. Levin et al. (2002). Given their very large cross-section dimension, the use of this panel technique is likely to solve the problems associated with the small-sample bias. However as noted by Gabaix and Ioannides (2004), their test does not correct for the potential autocorrelation in the residuals, which can severely bias the results regarding the unit root hypothesis and thus regarding the relevance of Gibrat’s Law.

Given the features of our data set, i.e. annual population data for the period 1925–1999 (except for the 5 years during the WWII period, 1940–1944), we argue that we can use standard unit root tests on individual city sizes.18 Given the decline of total urban population over the sample period we decided to allow the constant term in the growth rate, $K_0$, to be possibly trend-wise changing over time,19 i.e. $K_0=K_0t$ resulting in the following equation that will be estimated for each city separately:

$$\Delta \ln S_{it} = c_i + \left( \frac{c_i}{7} + \rho_i \Delta \ln S_{i,t-1} \right) + \sum_{j=1}^{p} \beta_{ij} \Delta \ln S_{i,t-j} + \epsilon_{it}$$

(10)

The first column in Table 8 shows the result of these city specific unit root tests. It shows the percentage of cities for which the null of a unit root is rejected at a 1%, 5% and 10% level.

For completeness, it also gives the outcome of two different panel unit root tests. The first is the earlier mentioned Levin et al. (2002) test, which tests the null of all series having a unit root versus the alternative of all series being stationary with the same autoregressive parameter. The second is the later developed Im et al. (2003) test that tests the null of a unit root in all series versus the alternative of some of the series being stationary (with a potentially varying autoregressive parameter) and some of the series being non-stationary. Hereby the latter test is thus somewhat less restrictive under the alternative.

Both the individual unit root tests and the Levin–Lin–Chu panel unit root test do not reject the null hypothesis of a unit root. Even at a 10% level the null is never rejected for any of the cities in the sample. The Im–Pesaran–Shin test on the other hand rejects the unit root null at the 5% level (it does not reject it at the 1% level), hereby providing some weaker evidence in favor of Gibrat’s Law. Overall, this would

---

17 The authors note this problem and propose the estimation of a restricted SUR model to overcome this problem, however the distributional properties of this SUR estimator are not known and given the small number of cities this is unlikely to solve the potential small sample bias.

18 As one of our referees pointed out, the power properties of the unit root test depend mostly on the time span of the data (here, 1925–1999) and not so much on the frequency of observation (see Campbell and Perron, 1991). In the German case, however, the large time span in combination with the annual data makes it possible to deal more adequately with the short-term dynamics (e.g. correcting for autocorrelation), which could be especially important surrounding large shocks like WWII, the division of Germany in 1948 or the re-unification with East-Germany in 1990.

19 Black and Henderson (2003) also include a deterministic trend in their estimated equation.
suggest overwhelming evidence in favor of Gibrat’s Law with all cities seemingly growing independent of their size. There is, however, one major caveat when drawing this conclusion from these standard unit root tests and that is (maybe not surprisingly) the WWII shock. As shown before, the large and sudden impact of WWII had a tremendous effect on the German urban landscape with large cities losing more population more systematically compared to the smaller cities in Germany. When performing a standard unit root test one implicitly assumes that the whole effect of this destruction during the war can be viewed as a one-time extreme realization from the distribution of the error term, i.e. $\varepsilon_t$ in (10). This however seems somewhat unlikely, instead WWII can be argued to have had a more substantial impact changing the deterministic components of city size growth, i.e. the constant and/or trend in Eq. (10). If this would be the case, to ignore it when performing a standard unit root test results in an underrejection of the unit root null hypothesis (see Perron, 1989; Perron, 1997). This would imply that Gibrat’s Law is potentially overaccepted by standard unit root tests in the case of German city sizes.

To allow for the possibility of a change in the deterministic components of city size growth, we follow Perron (1997) and estimate the following equation:

$$\Delta \ln S_i = c_i + \theta_1 DU_{it} + \theta_2 DT_{it} + \theta_3 D(T_0)_t + \rho \ln S_{i,t-1} + \sum_{j=1}^{g} \beta_j \Delta \ln S_{i,j} + \varepsilon_{it}$$

(11)

where $DU_{it} = 1(t>T_0)$, $DT_{it} = 1(t>T_0)$ and $D(T_0)_t = 1(t=T_0+1)$ and $T_0$ is the time at which the change occurs. The null hypothesis still remains that of a unit root, the alternative however changes from the series being stationary around a deterministic trend to the series being stationary around a deterministic trend that is allowed to change at time $T_0$. The exact timing of the break date is determined endogenously by the data (so that also $T_0$ is actually city specific), see Perron (1997) for details and Hansen (2001) for a discussion. We choose this procedure over the option of exogenously setting the break date at WWII to allow for the possibility that other events that could have had a major impact on German city sizes during our sample period like, for example, the separation from and subsequent reunion of West and East Germany (see Redding and Sturm, 2005, and Redding, Sturm and Wolf, 2007), which potentially may have left even a bigger mark on the evolution of some of the cities in our sample.

The results of these unit root tests when we allow for a one-time break are shown in the third column of Table 8. The impact of allowing for a one-time break in the deterministic components of city growth is quite striking. Instead of accepting the null of a unit root for all cities in our sample as the standard unit root tests did, now the unit root null is rejected for 74.2% of the cities in favor of these series being trend stationary with a one-time shift in this trend. The date at which the break occurs is almost exclusively found at WWII, which shows that WWII’s impact on city sizes is dominating that of other historical events affecting German cities during the second half of the 20th century.20 The high rejection rate of the unit root null hypothesis also implies that the relevance of Gibrat’s Law, which seemed evident based on the standard unit root tests gets a substantial blow. Correcting for the impact of WWII on the evolution of individual city size, Gibrat’s Law is found to hold only for about one quarter of the cities in our sample.

Together with the evidence obtained using non-parametric methods in the previous section, this dynamic evidence on Gibrat’s Law sheds substantial doubt on the relevance of random city size growth in case of Germany, especially for the post-WWII period.21 Instead the data seem to indicate that city growth does depend on initial city size, with smaller cities growing faster than larger

---

20 This is confirmed when doing the unit root tests on only the post-WWII period, the unit root hypothesis being accepted for about only 38% of the cities without allowing for a break in the series. When allowing for a break the division (1948) or reunification (1990) of West and East Germany did not show up as clearly as WWII in case of the whole sample period. Using a rather different analytical framework, Redding and Sturm (2005) do find evidence that the post-WWII division of Germany had a significant effect. Apart from the different framework, one reason that they find a stronger impact of this division on (west-) German cities is that their sample includes more (smaller) West-German cities located near the former border between West and East Germany. An interesting alternative method that could be useful in this case is the spatial interactions analysis as in Dobkins and Ioannides (2001).

21 The rejection of Gibrat’s Law could also be an explanation why we find mixed evidence on Zipf’s Law, see Fig. 3 plus discussion.
This faster growth of smaller cities does not comply with urban economic theories exhibiting random city growth. It suggests that other theories of urban growth are perhaps more relevant to explain the post-WWII experience of the German urban landscape. As the post-WWII and pre-WWII period are so different with respect to the implications regarding the relevance of urban economic theories, WWII seems to have had a crucial (initializing) role in the changing of both the type and evolution of the German urban system.

Having found compelling evidence on the irrelevance of random city growth as an explanation of the evolution of the German urban system, the next section, building on earlier work by Davis and Weinstein (2002) and Brakman et al. (2004), provides additional evidence based on the evolution of relative city sizes, i.e. the position of cities within the city size distribution, by which we try to distinguish between the two other competing theories of urban growth, namely increasing returns to scale and locational fundamentals.

6. Unit root testing and WWII’s impact on relative German city sizes

In previous work, Brakman et al. (2004) and Bosker et al. (2007), we already looked at the immediate impact of WWII on German relative city size. Drawing on the methodology developed by Davis and Weinstein (2002, 2004), these papers argue that the destruction during WWII was largely exogenous to the level of economic activity in cities and use the level of destruction during WWII as instruments for population growth during WWII when estimating the following equation:

\[ \Delta \ln S_{\text{post-WWII}} = \alpha \Delta \ln S_{\text{WWII}} + X \beta + \varepsilon \]

where \( S_i \) denotes city \( i \)'s size relative to total German population, i.e. \( S_i / S_{\text{tot}} \). \( X \) are other exogenous variables that can be included in the regression and \( \varepsilon \) is a random error term. Estimating this equation for Germany, Brakman et al. (2004) find evidence that the average German city had in 1963 returned to a relative city size of about 60% of its pre-WWII level. Bosker et al. (2007) extend this simple framework by estimating a threshold regression in the spirit of Hansen (2000) hereby allowing for the possibility of multiple equilibria and find evidence for the existence of two different equilibria, with the least destructed cities shifting to an equilibrium characterized by a larger relative city size.

The proposed framework, i.e. estimating Eq. (12),22 can however be argued to be subject to some caveats. First, the estimation results are sensitive to the choice of period over which post-WWII growth is calculated. Second, the estimation results are only able to describe the impact of WWII on the average German city, they are unable to say something about the individual experience of a particular city (the experience of the average city can even be argued to be of secondary importance when the fit of the regression is far from perfect). Third, and most important, the estimation of Eq. (12) is merely a static cross-section regression. Concluding that relative city sizes are mean reverting or random over time is impossible on the basis of such a simple cross-section. Instead, as argued by Hohenberg (2004) the historical evolution of the urban structure must always be studied in terms of fully dynamic models. This is exactly what we purport to do here. Exploiting the long time dimension of our data set, we are able to look at the evolution of each individual city’s relative size employing fully dynamic econometric estimation techniques.

Davis and Weinstein (2002) already mentioned the fact that the proper test for the persistence of shocks would be performing unit root tests on relative city shares. They refrain from doing this and instead resort to the static framework in Eq. (12) on the basis of the earlier mentioned low power of these unit root tests in small samples. Considering the extent of our data set, we think we are in a much better position to apply such unit root tests and can hereby provide much more dynamic evidence on the ‘mean reversion’ of relative city sizes. More specifically we estimate the following equation for each of the cities in our sample:

\[ \Delta \ln S_i = \xi_1 + \xi_2 \Delta \ln S_{i,t-1} + \sum_{j=1}^{P} \Delta \ln S_{i,t-j} + \eta_{it} \]

where \( S_i \) is the share of a particular city \( i \) in total German population, the lagged values of city growth included in the regression control for potential autocorrelation and \( \eta_{it} \) is a random error component. If \( \xi_1 \) is found to be significantly smaller than 0, city share is stationary around \( \xi_2 \) and any shock will not have a lasting effect. If on the other hand \( \xi_2 \) is found to be equal to 0 then all shocks are permanent and city \( i \)'s share in total German population follows a random walk.24 We estimate Eq. (13) applying Augmented Dickey Fuller tests to all of the cities in our sample. Table 9 shows the results; it also includes the results of the earlier mentioned Levin–Lin–Chu and Im–Pesaran–Shin panel unit root tests. The results of both the individual city share unit root tests and the panel unit root tests are at odds with the notion of city shares being stationary over time. The null hypothesis of a unit root in city share is not rejected for almost all cities in the sample.

22 The results of the unit root tests are much the same when looking at small or large cities separately. The rejection of the unit root null in case of the IPS test, seems to be largely due to the smaller cities in the sample however. Also the average autoregressive parameter of the smallest cities is somewhat smaller than for the largest cities, giving further evidence for a difference in growth process for small and large cities. Results are available upon request.

23 The same holds for the extended version allowing for multiple equilibria.

24 One may be concerned that city shares are by definition bounded, so how can they exhibit a unit root, which would imply an unbounded variance? Essentially, what we are looking at is whether the process, within the bounds, can be described by a random walk. Cavaliere (2005) shows that testing for a unit root in limited time series can still be done, but some adjustments have to be made to the test statistics. He also shows that, when the process is relatively far away from the bounds, so that the range constraints are rather loose, the standard unit root tests perform quite well. As this is the case in our sample (the maximum city share is about 0.07 and the minimum city share is about 0.001, both not very close to the bounds of 1 and 0 respectively), and considering that tests for ‘bounded unit roots’ that allow for a break in the deterministic component(s) are not readily available, we prefer using the standard unit root tests here. Note also that the presence of the bounds results in an overrejection of the unit root hypothesis (see Cavaliere, 2005). Given that we, when using the standard unit root tests, already find a unit root in virtually all series, using the bounded unit root test instead would probably not change our results.
This would constitute considerable evidence against the locational fundamentals theory. However, as was the case in our earlier unit root tests on city sizes (see Table 8), this conclusion does not take the possible different effect of the WWII shock into account. Also in case of city shares, the major impact of WWII could have resulted in a shift in the deterministic component of a city's relative size, i.e. a deterministic shift in the mean $\xi$. Such a change in the deterministic component could for example be due to a change in locational fundamentals as a result of the destruction in WWII.

To allow for this possibility we apply the following unit root test suggested by Perron and Vogelsang (1992) that allows for a one-time break in the mean of the series $\xi_i$ (endogenously determined by the data) and is based on the estimate of $\xi_i$ in the following regression:

$$\bar{s}_{it} = \xi_i \bar{s}_{i(t-1)} + u_{it}$$

where $u_{it}$ is the random error term and $\bar{s}_{it}$ are the residuals of a regression that projects $s_{it}$ on the deterministic component, i.e. a mean that is allowed to shift at time $T_b$. More formally:

$$s_{it} = \mu_i + \gamma_i DU_{it} + \eta_{it}$$

where $DU_{it} = 1$ if $t > T_b$ and 0 otherwise. Estimating $\zeta_i$ in this way controls for the possible one-time shift in the deterministic mean in the ‘first stage’ of the procedure (15) and estimates the autoregressive parameter, $\zeta_i$ in the ‘second stage’ (14). Perron (1990) called this the additive outlier (AO) model, which is appropriate to model a sudden one-time change, which is clearly the case when considering the destruction caused by the heavy bombardments during WWII. Perron and Vogelsang (1992) discuss the appropriate test statistics when testing for $\zeta_i = 1$.

The results of applying the AO-model to test for a unit root in German city shares under the null versus stationary city shares around a possibly shifting mean under the alternative are also shown in Table 9. As was the case for the unit root tests on city sizes, the effect of taking account of the possible special nature of the WWII shock (i.e. having an impact on the deterministic components of city shares) is quite substantial. At a 5% confidence level the unit root null hypothesis is rejected in favor of a stationary city share with a one-time break for 27% of the cities in our sample. Even more striking is the fact that for all the cities that are stationary, the timing of the break is (endogenously) found to be WWII. As in the unit root tests on city sizes, allowing for a one-time break in the deterministic component(s) the impact of WWII overshadows the effects of the other historic events (most noteworthy the separation from and unification with East-Germany) that could have had their impact on the evolution of individual cities and the city size distribution as a whole.

The evidence provided in Table 9 constitutes evidence against theories that fall under the locational fundamentals category. The ‘standard’ unit root tests reject stationarity of city shares for all cities in our sample except one (Hamm). When explicitly taking account of a possible shift in locational fundamentals during WWII, by allowing for a change in the deterministic mean around which a particular city is stationary, stationarity of city sizes is accepted for a much larger proportion of the cities in our sample. This does however not save the locational fundamentals theories as being relevant in the case of Germany. Still random shocks have a persistent effect on the relative city share of more than 70% of the cities in our sample. Furthermore, although random shocks are not persistent in the case of the cities that are found to be stationary, the extreme shock of WWII did have a lasting effect on the city share of those cities by changing the deterministic mean around which the city share is stationary. As the unit root tests only find the break in the deterministic mean but do not help us to understand why this break occurs (other than the date at which the break occurs), in Appendix B we tentatively look into some characteristics that distinguish the 17 “break stationary” cities (27% of our sample at the 5% level, see Table 9) from the other cities in the sample.

Overall, we think that this dynamic evidence on the effect of the large shock experienced during WWII constitutes considerable evidence against the relevance of locational fundamentals theories in explaining the evolution of the German urban landscape. The

\[25\] This in fact is corroborated in Bosker et al. (2007).
relevance of the other theories (random growth and increasing returns to scale) is somewhat harder to assess using the results of the unit root tests in this section.

7. Conclusions

Most of the empirical literature on city size distributions has focused on the USA. Other countries might experience a different evolution of their city size distribution, as this paper shows to be the case for West-Germany. Using a unique annual data set for 62 West-German cities that covers most of the 20th century, we look at the evolution of both the city size distribution as a whole and each individual city separately. The West-German case is of particular interest as its urban system has been the subject to some of history’s largest (exogenous) urban shocks, most notably WWII and the German division and subsequent reunification. Our data set allows for the identification of these shocks and provides evidence on the effects of these ‘quasi-natural experiments’ on the city size distribution as a whole as well as on each individual city separately that we subsequently use to distinguish between three competing theories that explain urban growth. Our first main finding is that (the evolution of) the German city size distribution is permanently affected by the World War II shock, more so than by any other shocks. Cities that have been hit relatively hard due to the substantial bombings and the subsequent allied invasion do not recover the loss in relative size. After the war, the city size distribution does not revert to its pre-WWII level, but shifts to one characterized by a more even distribution of population over the cities in the sample. Compared to the impact of WWII, the separation from and later reunion with East-Germany has had much less impact on relative city sizes. Our second main contribution is that we show that, once corrected for the heavy destruction during WWII, (panel) unit root tests that are used to test for the validity of proportional city growth reject Gibrat’s Law of proportional effect for about 75% of all cities. This constitutes considerable evidence against urban economic theories exhibiting random growth, a finding that is further confirmed by additional non-parametric evidence. Finally, evidence from (panel) unit root tests on relative city size also shed substantial doubt on the relevance of locational fundamental theories in explaining the development of the German urban system, even when allowing these fundamentals to change during WWII.

Overall our results are most consistent with theories emphasizing increasing returns to scale in city growth. However, we have not provided a proper test for this type of theory, concluding only in favor of it given the empirical evidence against the other two theories. Developing a proper test of urban growth theories exhibiting increasing returns is in our view a fruitful area of future research that could further substantiate our claim.

Appendix A. Data

Table A1 gives an overview of the 62 cities included in our sample.

<table>
<thead>
<tr>
<th>Table A1</th>
<th>West-German cities in our sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berlin West</td>
<td>Braunschweig</td>
</tr>
<tr>
<td>Hamburg</td>
<td>Mönchengladbach-Rheydt</td>
</tr>
<tr>
<td>München</td>
<td>Münster</td>
</tr>
<tr>
<td>Köln</td>
<td>Augsburg</td>
</tr>
<tr>
<td>Lübeck</td>
<td>Frankfurt am Main</td>
</tr>
<tr>
<td>Essen</td>
<td>Krefeld</td>
</tr>
<tr>
<td>Dortmund</td>
<td>Ludwigshafen am Rhein</td>
</tr>
<tr>
<td>Düsseldorf</td>
<td>Oberhausen</td>
</tr>
<tr>
<td>Stuttgart</td>
<td>Offenbach am Main</td>
</tr>
<tr>
<td>Bremen</td>
<td>Hagen</td>
</tr>
<tr>
<td>Duisburg</td>
<td>Kassel</td>
</tr>
<tr>
<td>Hannover</td>
<td>Freiburg im Breisgau</td>
</tr>
<tr>
<td>Nürnberg</td>
<td>Hamm</td>
</tr>
<tr>
<td>Bochum</td>
<td>Mainz</td>
</tr>
<tr>
<td>Wuppertal</td>
<td>Herne</td>
</tr>
<tr>
<td>Bielefeld</td>
<td>Mülheim an der Ruhr</td>
</tr>
</tbody>
</table>

These cities are included out of a total of 81 cities in the original data set\textsuperscript{26} on the basis of the availability of annual city population data over the period 1925–1999. The cities that were left out of the original data set were dropped on the basis of failing to comply to one (or more) of the following criteria:

I. More than two consecutive years with no population data;
II. Not able to correct for a so-called Gemeindereform, i.e. local government reorganization, that occurred in the early 1930s for several cities in the industrial Ruhr-area and for most of the sample cities during the 1970s.

\textsuperscript{26} For a detailed description of the original data set and its sources see Brakman et al. (2004).
The first exclusion criterion results in about 16 cities to be excluded from the data set. If at most two observations are missing we construct the city population for the missing years by interpolating (such a correction is made only 6 times). The other 3 are excluded based on the second criteria. Most cities in our data set are affected by the Gemeindereform of the 1970s and most cities in the Ruhr-area also by the Gemeindereform in the 1930s. In order to have the same unit of analysis in terms of city boundaries we have decided to take the city boundaries at the time of WWII as a point of reference. For example if due to a local government reorganization an adjacent town becomes part of one of the cities in our sample we extend this city boundary redefinition to all pre-WWII years if this redefinition happened before WWII and we ignore it if it happened after WWII. The fact that for most of the cities in our sample the exact number of people that is added due to a local government reorganization is recorded in the Statistical Yearbooks allows us to correct for this quite accurately.

More formally in case of a pre-WWII Gemeindereform at time \(T\) in city \(i\), we adapt the series as follows,

\[
\tilde{S}_{iT-k} = \frac{S_{iT-k}}{S_{iT-k}^{\text{new}}} \quad (16)
\]

where \(S_{iT}\) is the population at time \(T\) including the newly added towns, \(S_{iT-k}^{\text{new}}\) is the number of people living in the newly added towns, \(S_{iT-k}\) is the city population as reported in year \(T-k\), i.e. before the city boundary redefinition, and \(\tilde{S}_{iT-k}\) is the newly calculated, as if the new city boundary was already in affect, city population at time \(T-k\). If instead a city was subject to a post-WWII Gemeindereform this is incorporated as,

\[
\begin{align*}
\tilde{S}_{iT} &= S_{iT} - S_{i_{\text{new}}} \\
\tilde{S}_{iT+k} &= \frac{S_{iT+k}}{S_{iT}}
\end{align*} \quad (17)
\]

where \(S_{iT}\) and \(S_{i_{\text{new}}}\) are defined as above and \(\tilde{S}_{iT} (\tilde{S}_{iT+k})\) is the newly calculated, as if city boundary redefinition did not happen, city population at time \(T\) \((T+k)\). Thus the crucial assumption made in case of a pre-WWII Gemeindereform is that the city’s and its newly added town’s population grew at the same rate before the reform. Similarly in case of a post-WWII Gemeindereform the crucial assumption is that after the Gemeindereform the city’s and its newly added town’s population grew at the same rate.

**Appendix B. Stationary and non-stationary cities and their characteristics**

The unit root tests in Section 6 indicated that 17 cities experienced a one-time substantial impact of WWII on the share of their population in total German population. It turns out that 15 of these 17 cities experienced a negative impact of the WWII shock. To give some indication about possible differences between the break stationary and the non-stationary cities, Table B1 below shows the mean and standard deviation of several characteristics for 1) the 17 break stationary and 2) the 45 non-stationary cities. Also included are the results of a simple test of the equivalence of the means of the two groups given their variances and the final column contains this test when focusing only on the cities with a negative break in their city share. The tests for a significant difference in the mean of the characteristics show some interesting results.

<table>
<thead>
<tr>
<th>Table B1</th>
<th>Break stationary</th>
<th>Non-stationary</th>
<th>Negative break</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>War related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Housing destroyed</td>
<td>50.88</td>
<td>17.56</td>
<td>35.26</td>
</tr>
<tr>
<td>m³ Rubble per capita</td>
<td>20.46</td>
<td>6.56</td>
<td>9.81</td>
</tr>
<tr>
<td>Reconstruction aid</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>% Refugees 1960</td>
<td>0.15</td>
<td>0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>Geography related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to München (km)</td>
<td>366.24</td>
<td>128.96</td>
<td>419.38</td>
</tr>
<tr>
<td>Distance to Hamburg (km)</td>
<td>380.82</td>
<td>98.82</td>
<td>329.62</td>
</tr>
<tr>
<td>Distance to Köln (km)</td>
<td>155.08</td>
<td>93.61</td>
<td>182.87</td>
</tr>
<tr>
<td>Minimum distance to East-Germany (km)</td>
<td>178.83</td>
<td>65.48</td>
<td>182.96</td>
</tr>
<tr>
<td>City size related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-WWII population</td>
<td>369.250</td>
<td>640.162</td>
<td>241.725</td>
</tr>
<tr>
<td>Growth pre-WWII</td>
<td>-0.03</td>
<td>0.06</td>
<td>-0.02</td>
</tr>
<tr>
<td>Growth WWII</td>
<td>-0.29</td>
<td>0.20</td>
<td>-0.12</td>
</tr>
<tr>
<td>Growth post-WWII</td>
<td>0.13</td>
<td>0.20</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

Cities that have experienced a break in the deterministic mean of their city’s share in total German population have been hit significantly more severe during the war. The percentage of the housing stock destroyed is 15% higher and the amount of rubble per capita 11 m³ more than for cities with a random evolution of their city share. This shows that these cities have suffered
substantially more as a result of the bombardments during the war. Another interesting, possibly related, finding is that after the war these cities have had a lesser inflow of refugees than the cities with a non-stationary city share, possibly reflecting the fact that these refugees did not go to the more heavily destructed cities. The city size related characteristics confirm the notion that in those cities with a shift in relative city share more war damage resulted in a larger decline of this relative city share during the war. Mean WWII-growth is significantly less for these cities (~29% vs. ~12%). What is interesting is that although post-WWII growth is larger for these same cities, this higher growth does not compensate for the losses suffered during WWII, hereby resulting in a permanent impact on relative city share. Another interesting thing to notice is that pre-WWII city size (1939) does not differ significantly between the two groups of cities. Finally the cities are also compared on the basis of several geography-related characteristics. However distance from the later East-German border or to one of Germany's economic centers (Hamburg, München or the industrial Ruhr-area, to Köln) does not differ significantly between the two groups.

References

22 The final column of Table B1 indicates that these results are even more profound when focusing only on the negative city share breaks.