CESifo WORKING PAPERS

10435
2023

Original Version: May 2023
This Version: November 2023

What's in a Name? Initial Geography and German Urban Development

Duc A. Nguyen, Steven Brakman, Harry Garretsen, Tristan Kohl

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10435*

What's in a Name? Initial Geography and German Urban Development

Abstract

Place names or toponyms provide insight into the initial geographical characteristics of settlements. We present a unique dataset of 3,705 German toponyms that includes the date of the first historical record mentioning the settlement and the date that the settlement received city rights. Our findings are as follows. First, we show that the frequency of geographical toponyms as well as local geographical advantage reveal a city-size distribution adhering to Zipf's law. Second, we use the toponymical information to identify 168 geographical characteristics and empirically examine their importance for modern urban growth. We find that settlements with names referring to rivers, fords, churches, hills and historical clearing activities are associated with higher levels of population compared to places without any geographical characteristics as suggested by their name. Third, we document that the relevance of some of these characteristics for urban development changes over time. For instance, proximity to castles matters more for initial settlement growth than trade capabilities, highlighting the evolving significance of shifting from defensive geography towards water-based trade over time.

JEL-Codes: R110, R120, N900, N930, O100.

Keywords: toponyms, first-nature geography, Zipf's law, initial conditions, German urban development.

Duc A. Nguyen
Centre for International and Spatial Economics, Ruhr-University of Bochum / Germany
ducanhnguyen91@googlemail.com

Steven Brakman
Faculty of Economics and Business
University of Groningen / The Netherlands
s.brakman@rug.nl

Harry Garretsen
Faculty of Economics and Business
University of Groningen / The Netherlands
j.h.garretsen@rug.nl

*Tristan Kohl**
Faculty of Economics and Business
University of Groningen / The Netherlands
t.kohl@rug.nl

*corresponding author

This version: November 21, 2023

1. Introduction

Many cities in Europe have their origins in the Middle Ages, where the role of physical geography is often emphasized to have played a crucial role for initial urban development (e.g., Bosker and Buringh, 2017; Düben and Krause, 2023). Yet, relatively little is known about precisely which geographical circumstances, or so-called 'first-nature' characteristics, were crucial during the earliest development stages. Early geographical circumstances may still shape the modern urban landscape today even after becoming obsolete (e.g., Bleakley and Lin, 2012). This is an example of path dependence, and stresses the importance of initial conditions (see also e.g., Allen and Donaldson, 2020). Hence, knowing about the geographical circumstances that prevailed during the earliest days of a settlement may inform us about the relevant origins of today's cities. Our unique use of place names, or toponyms, enables us to go back to the early Middle Ages for a large sample of (mostly) German settlements. Crucially, our sample does not only include the (by now) largest cities but also more relatively smaller urban settlements. Recently, Düben and Krause (2023) showed for Ancient China that physical geography, like access to rivers, elevation or the climate more generally, is much more relevant for the growth of cities lower in the urban hierarchy.

We observe variation in (historical) first-nature geography and determine whether this initial variation corresponds with the observed variation in modern (early 20th century) city sizes. These early local characteristics survive in the toponym of a place, providing us with an indication of the geographical circumstances that were originally considered important from the very origins of a place's establishment. Toponyms convey the initial geographical features that span not only natural features that we might still observe today, but crucially also those that have disappeared naturally or through human intervention throughout history, as well as man-made ones such as castles, churches and monasteries that were around from the very beginning.[2]

We also account for other key historical characteristics, such as the first year that a place name was mentioned in historical records and the year the settlement received its city rights. City rights are important because it allowed places to organize their own markets, build walls, and to have their own courts and jurisdiction. Only larger settlements in feudal Europe were able to receive city rights as a developmental milestone; rights which were not granted in the agricultural feudal society outside the city. We create a novel proxy to measure the inherent early geographical advantage of a location around the time of its establishment through the

[2] The existing literature often distinguishes between first-nature and second-nature geography, with the former typically being assumed to be exogenous, while second-nature arises through interaction of people with one another (Redding 2010). In that sense, man-made characteristics are categorized as second-nature characteristics as a consequence of human interaction. Jedwab et al. (2020) observe a lack of man-made characteristics that are considered in the literature. This may be partly due to the inherent reverse causality with population growth, making it difficult to distinguish population growth from second-nature effects. Arguably, 'man-made' characteristics that are considered to be first-nature characteristics have the advantage that they are less susceptible to reverse causality.

'gap' measure which we define as the difference in years between when a settlement received city rights and when the settlement was first mentioned.

Our approach of inferring early geographical characteristics through toponyms differs fundamentally from previous studies that measure geographical characteristics as they are observed today.[3] Using current data as a proxy for the geographical circumstances of the past may be inappropriate due to depletion, destruction and other changes in the geography that happened throughout the many centuries. For example, if forests, fords or castles vanish over time, and if they were important drivers of early human settlement, we may not consider them simply because they are not physically observable anymore. Considering only today's geographical circumstances might thus disregard the fact that the surrounding geographical characteristics have changed over the centuries.[4] Furthermore, the location of the earliest settlement within a current (larger) city may have become blurred or unknown today, so that one may incorrectly infer a current characteristic as having been relevant very early on in a settlement's history. Our paper is not the first that assesses the relevance of first-nature geography (see, e.g., Combes et al., 2010, Ketterer and Rodríguez-Pose, 2018), but by focusing on the meaning of place names our approach is novel and it allows us to study a very rich menu of physical or first-nature geography features at the time of initial settlement.

Existing research in the economic literature is inclined to analyze the relevance of first-nature geography from mainly a transport-based and trade perspective by emphasizing the role of rivers, sea or road access of locations (see, e.g., Bosker et al., 2008 and Barjamovic et al., 2019). We considerably broaden the scope beyond these more standard transport/trade-based examples of water and land to also consider examples of elevation, vegetation, resources and structures. Here, we align ourselves with recent papers such as Düben and Krause (2023), Hanlon (2020) or Izdebski et al. (2020), however with a different and much broader menu of initial physical geography variables than merely the transport-based geography indicators to study urban development. All in all, our analysis of toponyms reveals 168 different sources of physical or first-nature geography.

We contribute to the measurement of historical urban development, where the variables may be early-history local population estimates, or some proxy for early economic activity, such as the number of archeological sites (Davis and Weinstein 2002, Bakker et al., 2021), environmental clues like pollen data (Izdebski et al., 2020), ancient logo-syllabic script records (Barjamovic et al., 2019), or ancient Roman tableware (Flückiger et al., 2019). However, with the exception of a few Roman-founded cities, almost all German settlements that still exist today came into existence only after 800 AD. Yet, economic data become sparser the further we go back into the past (Nagy, 2021). For example, Bairoch et al. (1988)'s dataset is widely used in the medieval urban economics literature (Hanlon and Heblich, 2020), but the

[3] While we are not the first to use place names as a data source – see, e.g., Oto-Peralías (2017), Lee and Lin (2018), and Villette and Purves (2019) – the distinguishing feature of our dataset is that it categorizes a settlement's toponym as interpreted by linguists.

[4] See, e.g., Combes et al. (2010) for an exception, as they include ecological information on soil quality in their analysis.

corresponding city population estimates mostly consider the larger European cities, which means that not just population data of cities at the early stage of development are often missing, but also smaller cities for which physical geography might matter more (Düben and Krause, 2023) are neglected.[5]

Our paper is also related to the literature in development economics on the role of initial, physical geography as a means to infer causality in a setting of natural experiments. Some studies looked at characteristics, such as terrain ruggedness and its impact on long-term development by inhibiting slave trade (Nunn and Puga, 2012), the impact of air pollution from coal burning on British city growth (Hanlon, 2020) or the impact of abnormal floods (Chaney, 2013), the Tsetse fly on population density (Alsan, 2015), and the economic impact of droughts (Hornbeck, 2012). Compared to these studies, toponyms provide us with a unique and very long-term perspective on economic development.

Lastly, we contribute to the literature on the important regularity in urban economics that concerns the city size distribution of cities and whether or not it follows a Zipf distribution. Zipf's law holds in many countries using present-day data, including Germany (Giesen and Südekum, 2011; Ioannides and Overman, 2003; Soo, 2005; Nitsch, 2005; Dittmar 2011).[6] Finding the same empirical regularity within a measure of geographical quality as well would indicate a more intimate relationship between geography and city size as hypothesized by Krugman (1996) and Gabaix and Ioannides (2004). However, this relationship has so far not been explored empirically likely due to the difficulty of separating geographical advantages from second-nature geography (Redding, 2010). With our dataset, we arguably find first empirical evidence for this relationship.

Our main results are as follows. First, we find that toponym frequency and early geographical advantage (defined by the time gap between city rights received and first recorded mention of the place) both exhibit a Zipf distribution. These results support the strand of the literature that explains Zipf's law for cities through variations in early geographical advantage (Krugman, 1996; Gabaix and Ioannides, 2004). Second, some early settlement characteristics explain modern population even after many centuries. We find that for frequently observed characteristics, rivers, fords, river mouths, early churches, and historical clearing activities are notable contributors. Also, the proximity to a castle or a lake mattered. Together, the results indicate that water-based trade (characteristics) became more important over time, consistent with some previous findings (Bosker et al., 2013). Finally, the results indicate a heterogeneous interaction of second-nature geography with

[5] The data are also often based on educated guesses, which are prone to errors (Bairoch et al., 1988, pp. 298). Only after the introduction of regular censuses at the beginning of the 19th century is the population in cities systematically and centrally recorded and thus reliable. For example, census city level population data became available for France, the U.K., and the U.S. around 1800 and later, while they became available only later for Germany.

[6] This distribution has generally been generated through some form of Gibrat's law of proportionate growth (Gabaix, 1999a,b) within some urban growth models (Gabaix, 1999b; Eeckhout, 2004; Duranton, 2006; Rossi-Hansberg and Wright, 2007). However, as Gibrat's law by itself has no natural economic explanation, an economic explanation to Gibrat's law is typically imposed. This in turn may add too much freedom in the choice of the economic mechanism that governs Gibrat's law.

different first-nature geography. For instance, for characteristics associated with water(-transportation) in particular, we find that rivers, floodplains and sources benefitted from being an early, pre-industrial city, while we find the opposite case for fords or river mouths.

The remainder of this paper is organized as follows. The motivation for the underlying primary data source is discussed in more detail in section 2, where we will elaborate on the purpose and usefulness of toponyms. Section 3 shows how we interpreted the original data to create our own dataset, and we will elaborate on details of the dataset itself. As a way to characterize the resulting city-size distribution, section 4 presents the results on the Zipf distribution. Section 5 presents our main results by estimating the impact of geographical characteristics on local economic activity. Section 6 concludes. The Data Appendix presents details on the main data sources used and provides examples of toponym changes throughout history, as well as details on the first recorded mention of a settlement and on the year a settlement received city rights.

2. Related literature

We use toponyms to identify *initial* geographical characteristics, which distinguishes our approach from the existing literature that mostly considers geographical characteristics that are observed *today*. Although rivers and mountains do not change much over time, other characteristics such as forests or fords may have disappeared by forces of nature or human intervention. Only focusing on currently observed characteristics ignores these aspects and may underestimate the role of historical first-nature geography which may have disappeared, and possibly misinterpret how second and first-nature geography interact. Toponyms are an ideal source for these initial geographical characteristics: during the Middle Ages, illiteracy reigned which necessitated accurate toponyms reflecting the immediate geographical surroundings of a settlement as a means of orientation and navigation for the majority of the population. The role of toponyms for the illiterate will be elaborated in more detail in section 3.1. Furthermore, to measure the role of geography on the path dependence of settlements (see Bleakley and Lin, 2012, Michaels and Rauch, 2018), we would ideally measure the geographical conditions during the early days of a settlement. As these initial geographical conditions are preserved in our approach through its initial toponym, they are resistant to either changes in geography or place names over the many centuries, thus allowing us to measure the influence of these initial characteristics even until today.

A variety of physical characteristics and its role on urban development had been investigated so far in the literature. The role of ports, for example, has been analyzed by Acemoglu et al. (2005) and Bosker et al. (2008), and portage sites by Bleakley and Lin (2012). The importance of navigable rivers or sea-access is part of the analysis in Düben and Krause (2023), Bosker and Buringh (2017), Henderson et al. (2017), and Rappaport and Sachs (2003). Armenter et al. (2014) studied at the effect of bridges on local economic growth. It is important to note that first-nature geography can have manifest itself through other, indirect channels. For example, Alsan (2015) explains the low population density in Africa by the prevalence of the tsetse fly in certain areas. This fly transmits a parasite that causes sleeping

sickness in humans and nagana in domesticated animals. In areas where the fly was abundant, technological advances in agriculture lagged behind and resulted in lower population density and fewer urban centers; as well as no development of plough agriculture which more likely led to slaves being used in the local workforce. Nunn and Puga (2012) show that terrain ruggedness in general increases barriers to trade because it is more difficult to build roads, bridges, and railroads. Within Africa however they observe that greater ruggedness as a barrier to (slave) trade is correlated with higher income instead of lower income. Similarly, Dalton and Leung (2014) find that mostly males were enslaved in the transatlantic slave trade and shipped to the Americas. The relative shortage of males to some extent could explain why polygamy is more widespread in West Africa as compared to East Africa, as the Indian Ocean slave trade did not have a male bias.

The role of climate is for example explored by Fenske and Kala (2015), who observe that the cooler temperatures near slave trading ports stimulated slave exports from those ports. Similarly, Chaney (2013) shows that abnormal floods resulted in higher food prices, more social unrest, fewer changes in religious leadership and more construction of religious buildings. This reallocation of resources to religious structures from more productive uses in turn can negatively impact economic development.

Finally, Hornbeck (2012) studies the long-term economic effects of the 1930s drought in the U.S. known as the Dust Bowl, in which strong winds in combination with a long period of droughts eroded the topsoil of the American Plains. Land prices fell, stimulated out-migration and diverted immigration. The migrants re-located towards coastal areas which became denser. These demographic consequences still endure until today.

Table 1 presents a selection of related studies on the direct role of geographical characteristics, be it natural or man-made, on local economic activity. Most characteristics considered in previous studies are easily observable ones, such as river access, sea access or altitude. Also, the emphasis in the literature is on characteristics that seemingly benefit trade. Nevertheless, and key to our motivation to use toponyms, is that the variety of characteristics considered in the literature so far is rather limited as compared to the variety of nature. The use of toponyms will therefore not only introduce some not so obvious characteristics, such as hedges, floodplains or clearings, it will also add considerably more distinctions within certain characteristics. For instance, we do not only consider the typical river, but also fords, river islands, river mouths, floodplains and streams. The same holds for altitude, where we can distinguish between mountains, stone, hills or simple 'elevations'. In essence, due to the large variety of meaning in toponyms, and by implication a large variety of geographical characteristics, we can compare geographical characteristics in a more disaggregated manner, and also substantially expand on the range of first-nature geographic characteristics beyond the current scope of the existing literature.

<< TABLE 1 ABOUT HERE >>

3. Data

This section describes the main dataset for our empirical analyses. We first discuss the purpose of toponyms during the Middle Ages. Second and most importantly, we introduce our toponymical dataset, which we will use to obtain our indicators of geographic characteristics in original place names. Third, we explain how we obtained information about two key, historical variables, i.e., the year a place was first mentioned in any record, and the year in which a place was awarded city rights. Finally, we briefly discuss the 1910 population data which are used as one of our main outcome variables, along with other variables included in the dataset.

3.1. The initial purpose of toponyms

During the Middle Ages, the vast majority of people were illiterate. Toponyms fulfill the crucial part of conveying the geography of the place of interest without the need to read. They acted as orientation points between places and within the landscape. That is, toponyms were developed out of a need to describe the immediate physical surroundings. The need for this orientation becomes apparent in the case of nearby settlements that originally shared the same name because of similar local geographic characteristics. In the context of Germanic place names, it is common to add a distinguishing prefix to the place name, typically referring to the relative age (old, new) or position of the place (north, south, high, low), thus creating local individualization of the toponym, a phenomenon described as *toponymische Raumorganisation* ('toponymical space organization', Brendler, 2008).

Even though there is no official convention in the naming of a settlement, it should be noted that 'fantastical' names that do not describe the geographic characteristics at all, such as Friedrichsgabe (present of Friedrich), Glückstadt (lucky place) or Freudenberg (joy castle) were not common at all. Also, when new names were needed after places were merged or annexed – a process which mostly occurred in the 20th century – a typical suffix is the rather encompassing -tal (valley), suggesting that an accurate description of the local geography through toponyms became less important when the majority of the population had become literate.

It is for this reason that the study of toponyms in our context is best done in the Old World. For instance, it is relatively common to find toponyms in the United States that refer to cities in the Old World, such as New York or New Orleans, or honorary names such as Bismarck, or the many Washingtons or Lafayettes. This signals that accurate geographical descriptions embodied in the toponym were far less needed with the adoption of widespread, sophisticated cartography around the time of settlement. Second, planned cities and as such, names chosen by a planner or ruler, were not common in the Middle Ages when the vast majority of settlements were first mentioned in some record. As such, we consider the majority of the names that mention some geographical characteristics to accurately reflect the local geography.

3.2. Toponym source

Our toponyms are taken from Niemeyer's (2012) *Deutsches Ortsnamenbuch*, hereafter referred to as 'the book'. The book is a compendium containing the etymological development and the interpretation of toponyms for places with a present-day population of above about 7,500 inhabitants in the German cultural sphere of influence, i.e., the German Reich right up to World War II, as well as Alsace-Lorraine in France and the German part of Switzerland.[7] Ultimately, we end up with 3,705 places in our sample.[8] **Figure 1** displays the geographic distribution of the places in our dataset, along with an indication of their city-rights status.

<< FIGURE 1 ABOUT HERE >>

A toponym usually consists of a prefix and a suffix. Due to the structure of the German language, the suffix usually conveys the main characteristic, while sometimes a prefix further describes this particular characteristic in more detail.[9] For example, Gladbach has the suffix -*bach* describing a stream, and the prefix *Glad*- describing that the stream is 'smooth', typically referring to a slow flow. For our purposes, we mostly rely on the information contained in the suffix to identify the characteristic. That is, for Gladbach, we ignore the 'smooth' part in 'smooth stream'.

Note that a settlement's toponym may contain several characteristics which can also be a combination of a geographic and non-geographic meaning. For example, Querfurt translates as 'mill castle' and therefore contains two characteristics related to structures. Düsseldorf translates as 'village on the river Düssel', so it contains the non-geographic 'village' and the geographic 'river'. Overall, 2,110 entries (56.9 per cent) only contain a geographic meaning, 1,088 entries (29.4 per cent) only contain a non-geographic meaning and 287 places (7.7 per cent) have toponyms containing both a geographic and non-geographic meaning. For 220 entries (5.9 per cent), the linguists in the book could not provide a preferred or reach consensus on the interpretation of the toponym.

We aggregate suffixes that share the same geographical meaning into individual characteristics. For example, toponyms containing the Celtic *-dunum*, *-stein* ('stone'), -*berg*/*burg* ('castle'), and the Slavic -*grad/grod* ('castle', e.g., Graz, Belgrad) are all considered as meaning 'castles'. For the sole purpose of retaining an overview, we further categorize these individual characteristics into 8 overarching categories – Water, Structures, Vegetation,

[7] See Appendix A.1 for detailed examples of how entries in the book were coded into our dataset.

[8] Our dataset initially includes 3,738 useable toponyms, yet we exclude 15 places which the book refers to as planned cities, or which we suspect to be planned given the very small differences between the year when these places were first mentioned and the year when these places were awarded city rights. Another 17 places are omitted because these merged prior to 1910 and would otherwise introduce measurement error in our dependent variable, which is population measured for places existing in 1910. For a similar reason, we omit 1 place which was split into multiple smaller places prior to 1910. The remaining 3,705 places are the basis for our empirical analyses.

[9] This is usually not the case for old, Roman cities, such as Köln (colony) or Bonn, which has no such prefix or suffix. We consider these to be standalone suffix toponyms without a prefix.

Elevation, Land, and Resources, and Miscellaneous, spanning 168 different geographical characteristics – and 1 No-Geography category. **Table A1 in Appendix A** describes these characteristics with their basic word (typically a suffix) in the second column. However, many of these individual characteristics are observed only a few times. For example, we often observe toponyms referring to rivers and streams, but references to islands and fjords are much less common.

<< TABLE 2 ABOUT HERE >>

Table 2 presents non-unique geographic characteristics in our sample, i.e., with more than one observation. There were several cases with settlements which experienced a change in the meaning over time (163 cases, 4.40 per cent of all cases). Since we are interested in the initial characteristic, we importantly only consider the original meaning of the first record of a place's name. For example, Münster was initially recorded as 'Mimigernaford' that refers to a ford of the people of Mimigern. Over time, it became known for its 'Münster' (monastery) rather than its ford. Apart from our focus on the earliest geographic circumstances, relying on the original meaning also avoids potential mistakes introduced by changes in languages and transcription errors that accumulate over time, as well as potential simultaneity bias of man-made characteristics.[10]

A second challenge is that a settlement may be subject to multiple interpretations (182 cases, 4.91 per cent of the sample). For example, Steinhagen translates as 'stone hedge' or 'mountain hedge', where we are only sure about the 'hedge' part and assign this settlement accordingly to the hedge characteristic. In such cases, the book usually provides further guidance as to the preferred or consensus interpretation, and we do additional desk research to verify the correctness of this interpretation based on a place's known history and geography. These cases are excluded in later robustness exercises. Finally, some settlements have either no preferred interpretation between multiple interpretations, or there are simply no interpretations available (220 cases, 5.94 percent of the sample). In this case, we do not assign the toponym to any characteristic and we later also exclude these cases in the robustness exercises.

3.3. First-record year and city-rights year

The first recorded mention of a place is not an official 'foundation' date, such as the date of incorporation of places in the United States Census Bureau definition of municipalities. Instead, it should be considered as the first time the place was of sufficient interest or importance to be recorded in writing. Hence, with the few exceptions of planned places and cities such as Karlsruhe, the first recorded mention does not necessarily correspond to the date of foundation or the date of first settlement. Typically, the place itself is mentioned in a manuscript or document along with the date given. Occasionally, not the place itself is

[10] For a detailed discussion of how we deal with names that change over time, see Appendix A.3.

mentioned, but a person named after the place, such as 'Wernherus de Menkemere' from Maikammer (1260) or 'Waltherus de Aken' from Aken (Aachen) (1227). So, while the first recorded mention is not necessarily the year the first inhabitants populated the location, it is arguably the closest possible estimate available.

The first recorded mention year is retrieved from the book. We cross-referenced this year with other sources such as the place's official online website. We assign the first recorded mention "around 900" as simply 900. "15th century" is coded as 1450, unless we find additional sources with more precise information. Overall, these imprecise dates for first-record years only affects 149 (4.02 per cent of) observations and are excluded in robustness checks.

City rights allowed places to organize their own markets, build walls, and have their own courts and jurisdiction. We think of the *Magdeburger Stadtrecht* and the *Lübecker Stadtrecht* as the prototypical city rights. The date of receiving city rights is also retrieved mostly from the book, which typically draws on various official sources and which we cross-check with other sources. Yet, some of the original documents may not have survived throughout history. In these cases, one has to examine documents that reconfirm city rights at a later stage. The city-rights year can also be approximated by the year in which a place's walls or courts were first mentioned, or walls were depicted in the place's coat of arms or seal.

We also assume that places have de facto city rights if they are named as an *oppidum* (lat. fortified place) or a Wigbold (limited city rights) at some point in time. A place may receive 'proper' city rights after it was named an *oppidum* only a few years later, but being named an *oppidum* signals that the place already resembled a city with city rights, or was at least sufficiently large to be one. Nevertheless, we should also note here that an oppidum is not a clear-cut city right in itself, and that a place with 'proper' city rights may not even build walls. Also, very rarely, some places may never receive or apply for city rights, even if they are considered as a city by modern standards. For example, Offenbach, which in 1910 already had a population exceeding 75,000 and is by all means considered a city today, has never possessed anything resembling city rights.

Figure 2 shows the distribution of first recorded mentions for our sample of 3,705 places. While most of the city rights were issued before the 16th century, another spike in city rights is visible spanning the mid-19th to late-20th centuries. The latter reflects population growth after the advent of industrialization, and the merging with other places after a series of *Gebietsreformen* in the 1960s, which led to many municipalities applying for *symbolic* city rights to reflect their size as a de facto city.

<< FIGURE 2 ABOUT HERE >>

Figure 3, panel (a) shows the kernel estimates of the first recorded mention of the 5 most common characteristics with a geographic meaning – stream, castle, river, mountain, and tree – as well as a selection of toponyms without geographic meaning in panel (b). We find a bimodal distribution especially for toponyms without a geographic meaning, streams and rivers; as well as castles, mountains and trees. The likely explanation for the two peaks can be found in the Frankish colonization, which took place in the three centuries following the Battle of Zülpich in 496AD. It seems that Frankish colonization settlements primarily have names without geographic meaning, or have the names of streams and rivers. The kernel estimates in panel (b) illustrate this in a sense that typical Frankish colonization settlements without geographic meaning, such as *-heim* or *-hausen*, are first mentioned around that time. Also notice the Gallo-Roman *-acum* suffix ('place of'), which occurs before 500AD, and the *-dorf* suffix, which is not associated with the Frankish colonization.

<< FIGURE 3 ABOUT HERE >>

3.4. Population and urban potential

To measure modern urban development, we consider the 1910 population census data as the size of these German places in 'modern' times, and not today's population data. The single most important reason for this decision is the widespread mergers of German places in the course of the 20th century, which was a relatively rare occurrence until 1910. A consequence of this merger wave is the lack of population data today on very small, already merged places. That is, the 1910 population census data are mostly complete compared to today's population data in terms of their coverage of places in our sample. The population data are from the public website *Gemeindeverzeichnis.de*, which compiles regional population census data at the municipality level within the 1910 borders of the German Reich.[11]

For the regressions that are introduced in section 5, we will use the available 1910 population data to construct a simple measure of urban potential (UP) that will serve as control variable, with

$$UP_i = \sum_j^n \frac{pop_{i,1910}}{\tau_{i,j}} \quad (1)$$

and where we follow Bosker et al. (2008, 2013) in setting distance to linearly affect urban potential. However, we simplify the measure by not giving different distance weights depending on the implied mode of transportation to reflect estimated differences in medieval transport costs. One reason is the emergence of railroads that made such discounting less relevant in 1910. That is, $pop_{i,1910}$ is the population in 1910, while τ is the simple Euclidean distance measure in kilometers from place i to place j. These bilateral distances are computed based on the places' geographic coordinates (i.e., degrees latitude and longitude) obtained

[11] The list of individual 1910 censuses can be retrieved from www.gemeindeverzeichnis.de/gem1900/gem1900.htm?quellen/quellen.htm. The 1910 population data can be readily retrieved from the respective statistical offices for places in Austria, Belgium, the Czech Republic, Luxembourg, and Switzerland.

with the help of Google Maps. **Table 3** provides the descriptive statistics for the data pertaining to population, urban potential, first-record year and year in which city rights were awarded.

<< TABLE 3 ABOUT HERE >>

4. Zipf distribution of first-nature geography

If first nature variation is able to explain overall city size distribution, its own distribution should be consistent with a rank-size curve or even Zipf' law, which is a special case of the rank-size curve.[12] Such a relationship would provide an additional understanding of the potential validity of Zipf's law and show the importance of initial conditions for the law to hold. Our dataset allows us to consider the relationship between initial first-nature geography and Zipf's law. In section 4.1, we look at whether the frequency of geographical characteristics as suggested by the toponyms follows a Zipf distribution. In section 4.2, we introduce a proxy measure for early geographical advantage, defined by the 'year gap' between the year city rights were received and the year of the first recorded mention of a settlement, and ask whether this proxy follows a Zipf distribution.

4.1. Frequency distribution of geographical characteristics

With toponyms, we look at early settlement patterns near some specific geographical characteristics, capturing the choice of the earliest settlers. One advantage of using toponyms is that assigning local characteristics to settlements is straightforward, as we do not have to choose a maximum distance from the settlement to some geographical characteristic. Another advantage of using toponyms is that the implied characteristics are measured in a binary and thus standardized manner, which simplifies quantification. Otherwise, we may have to choose, for example, around which height a mountain can be considered a mountain and we would have to make similar choices for all characteristics. Overall, toponyms allow us to consider the frequency distribution of a large variety of geographical characteristics in our sample.[13]

Following Gabaix (1999a), the probability of the frequency $freq_l$ of a geographic characteristic l being larger than some F is equal to

$$\Pr(freq_l > F) = \frac{k}{F^{\zeta}} \qquad (2)$$

[12] Gabaix (1999a,b) provides an explanation why city size distributions may behave according to Zipf's law. If city growth follows Gibrat's law with a common mean and variance, then a Zipf distribution follows, at least in the upper tail. Krugman (1996) conjectured that if we assume that the variation of a static landscape is random, it implies a Zipf distribution in the upper tail.

[13] Villette and Purves (2018) show a Zipf distribution of micro-toponyms, i.e., names given to natural features in the canton of St. Gallen. Their study differs from ours in terms of geographic coverage and use of toponyms; ours uses the actual interpretation of the earliest known name by linguists, while theirs is based on a lexicon of typical toponyms that is subject to misinterpretations discussed in Appendix A.3.

for some positive constant k and ζ. We regress the following through ordinary least squares (OLS),

$$\log\left(rank(freq_l) - \frac{1}{2}\right) = \alpha - \zeta \log(freq_l) \quad (3)$$

for some constant α. When the negative slope ζ being estimated is close to 1 and with a straight line, i.e, high R^2, the frequency of geographic characteristics follows a Zipf distribution. As is standard in the literature, the rank is adjusted by -0.5 to correct for potential small sample bias (Gabaix and Ibragimov, 2011).

The results are reported in **Table 4**. In panel (A), we only consider places with geographic meanings. We approach a Zipf coefficient with ζ approaching 1 as we increase the minimum frequency of a given geographic characteristic. As noted by Brakman et al. (1999), the negative correlation between rank and frequency at the upper tail may break down for a lower frequency. In our case, a possible explanation is that integers may be an issue for lower frequencies, which, combined with potentially many unique names that do not show up in the sample, lead to lower coefficient estimates. **Figure 4** shows the Zipf plot for the specification of **Table 4**, panel (A-I), with no minimum frequency. In **Table 4**, panel (B-I), we further restrict the sample to all geographic toponyms except Structures. Again, we approach a Zipf distribution with increases in the minimum frequency. Finally, as the book may be biased towards interpretations of places of present-day Germany, we repeat our analyses for a sample of places within the current borders of Germany (panels B-II and C-II), with little change to the estimated coefficients.

Based on these findings, we conclude that settlement behavior around geographical characteristics exhibits a Zipf distribution in the upper tail. However, although these results are intriguing and support of the idea that Zipf's law for cities may have its origins in first nature-geography, as suggested by Krugman (1996), more evidence is warranted, ideally by measuring an individual location's geographical advantage or land quality early in time.

<< TABLE 4 ABOUT HERE >>

<< FIGURE 4 ABOUT HERE >>

4.2. Distribution of early geographical advantage

One shortcoming of the above exercise that is more in line with the evidence of Krugman (1996) is that it only addresses that the frequency of settlements across geographical characteristics follows a Zipf distribution. It does not directly link the quality of local characteristics to a city size distribution. The inherent geographical advantage of a location can be seen as an index of geographical characteristics, which determines a location's economic activity (see Gabaix and Ioannides, 2004). Instead, we measure this implied index of geographical advantage itself through the gap in time between the year that city rights were awarded, and the year the city was first mentioned in any record.

If the time between the first mentioning of a city and the year in which city rights were granted is small, the place has undergone rapid development in its early days and we interpret this is an indication of favorable first-nature characteristics. First, this link should be relatively stronger in the absence of increasing returns effects or when first-nature geography is a relatively stronger determinant of local economic activity, i.e., when population levels are not very high. This would be the case when a location was initially settled, as there was little human interaction or buildup of endogenous amenities early on.[14]

However, we can think of some characteristics to be predestined as natural second-nature places early on. For example, we may think of choke points such as fords, bridges and hedges as a natural hub for merchants, where trade facilities developed relatively early on. Following Duranton and Puga (2004), increasing returns effects would then come from the sharing of indivisible facilities, such as markets where supply and demand can physically meet.[15] Markets could be organized after obtaining market or city rights in the Middle Ages.

Although bartering also happened in rural settlements without those rights, a central market place was only allowed if the settlement possessed the right to hold markets as granted by the ruler in the feudal system. The emphasis here is not that bartering did not happen outside of markets, but that markets, as an early urban indivisibility, were an initial source of second-nature geography after market or city rights were granted.

The book also mentions a place as a city or a market, indicating the respective rights received in the past. However, since a place may have received market rights before city rights, or entirely skipped these market rights to obtain city rights directly, we do not consistently know whether a place had market rights before city rights. The relatively low number of places that the book has marked as having had market rights at some point in history, 275 (7.4 per cent), indicates that city rights data, with 2,093 entries (56.5 per cent) is more readily available.

The years between the first recorded mention up to the point that city rights are obtained can then be considered a time where important indivisible facilities such as a marketplace, criminal courts or walls did not yet exist, thus restricting the second-nature impact through sharing and matching. Another source may come from the sharing of a relatively wider variety of intermediate inputs that were imported. We could measure this through the variety of guilds that were present in a city. For example, Lübeck as the de facto capital of the Hanseatic

[14] Even around the year 1500 and within the Holy Roman Empire, around 90 to 95 percent of cities were below a population of 2,000. Furthermore, 85 percent of these cities were considered dwarf cities with a population below 1,000 (Schilling, 2004, p. 8). For these dwarf cities, the transition between rural and urban area is considered 'smooth' (Stoob 1985, p. 152). Within the literature, Bosker et al. (2008) and Bosker et al. (2013) showed both the impact of first- and second-nature geography on economic activity throughout time. While the first study on Italian cities above 10,000 between 1300AD to 1861AD finds no significant impact of second-nature geography, the latter does for cities larger than 10,000 in Western Europe, North Africa and the Middle East. However, our dataset includes much smaller dwarf cities as well, which is reflected by the larger number of cities that we consider around Germany. For example, the number of cities in the Bairoch et al. (1988) database only includes 244 cities within the current German border, compared to 1,809 cities in our dataset.

[15] See Konishi (2000), Berliant and Konishi (2000) on the theoretical literature of market places. However, they did not explore the relevance of market places empirically.

League and had around 36 different guilds at the beginning of the 15th century, while Braunschweig had around 13 in year 1445 (Tuckermann, 1906), despite their roughly equal population of 25,000 and 18,000 respectively in 1400 (Bairoch et al., 1988). However, we again consider these sources to be mostly relevant to the largest cities after receiving city rights, whereas the majority of cities in our dataset would still be classified as dwarf cities.[16]

In any case, we will proxy the inherent, inverse early geographical advantage by the difference of the year when city rights were first issued and the year of the first recorded mention for place i in country c, that is:

$$\log(\text{gap}_{ic}) \equiv \log(\text{cityrights}_{ic}\text{-firstrecord}_{ic}) \quad (4)$$

A small gap corresponds to a large inherent advantage very early in its settlement, and vice versa.

The advantage to a 'traditional' historical economic activity measure, such as the population estimates by Bairoch et al. (1988), is that our proxy allows us to consider all cities in the dataset and it is not limited by the availability of data in a given year. A quite similar rationale to our measure was made by Ioannides and Overman (2004), who indirectly measured variations in first-nature geography through the age of places in the United States, as they reasoned that places that were settled earlier were better places in terms of first-nature characteristics.

On the other hand, our main rationale is that good places in terms of first-nature were settled faster from the point of a primitive rural settlement, measured by first recorded mention, until some threshold was reached where it was considered a city de jure. Arguably, this would be a more direct measure for a location's strength, as we consider the first-nature characteristics around the time of first settlement with some threshold point.

Also note that there was no fixed threshold point, as these rights were granted by the feudal rulers of the land. City rights were issued in two waves (see **Figure 2**): the initial wave started in the High Middle Ages after 1150 until around 1700, and city rights were granted in exchange for a fee, taxation, or provision of military support among other services to the feudal rulers (the King, Fürsten or bishops) of the principalities of the Holy Roman Empire. These city rights tend to encompass the right to hold markets, to fortify, to create courts, or for staple rights that required traders to hold their products in warehouses in order to first be sold to citizens.

As noted before, our data do not show that planned cities that were founded by some ruler were common.[17] Instead, the vast majority of cities existed as recorded places before they were issued city rights and as we will see in this section, the time 'gap' between first

[16] In a similar line of argument stemming from intermediate inputs as well as from matching and learning, there were designated streets for some professions, as indicated by its street name. In the city of Lübeck for instance, bell founders were found around the Glockengießerstraße, butchers around the Fleischhauerstraße, and hucksters around the Hüxstraße. This may be of interest for future research in a similar vein.

[17] We have noted 15 planned cities in our dataset.

record and city right is often large. Thus, the notion that cities were planned around some geographical characteristic that was perceived as beneficial by the ruler is empirically unfounded.

Instead, following the historian Irsigler (2015), to allow for growth of the economy within their territory, the ruler gave city rights to communities with some pre-urban or urban characteristics. This elevated these communities to a better position as compared to the neighboring communities that still operate under the feudal system. In contrast to the free citizens of a city, who enjoyed the right of inheritance or to own real estate among other rights, the majority of people outside the city were serfs that worked the land of their lord or had to provide other forms of services/produce, primarily in exchange for the security provided by said lords. The saying "*Stadtluft macht frei*" (city air makes you free) reflects the custom that serfs were free after a year of not being claimed by their feudal lord, which also made cities islands of asylum for serfs that wanted to escape serfdom. That is unlike today, where we may call a place a 'city' if its population was over some strict threshold, no uniform criteria existed to make a place eligible to be awarded city rights. Instead, cities typically sprung out of already mature communities of varying size, supporting the idea that the gap measure indicates for most places the time until it reaches some developmental stage instead of, with very few exceptions, being founded as cities out of nowhere.

Yet, we cannot rule out that rulers awarded city rights preferably to places with higher expected economic growth conditional on becoming a city, such as places along trade routes that would be ideal for holding markets. The average population requirements to be awarded city rights may then be relatively lower for these places. However, this works for our purposes as the gap measure would be closer in line as a proxy for geographical advantages à la Krugman (1996): Information about the city rights awarded not only convey those already well-developed communities which received city rights through being large enough; they also convey information about those relatively not so well-developed communities with relatively high expected growth rates only once they have their underlying geographical advantages 'unlocked' through the award of city rights.

So far, we have described city rights as the de jure transition point from which on important indivisibilities were allowed, but also other agglomeration economies à la Duranton and Puga (2004) as a source of second-nature geography may take place. For example, a resulting increase in urban population and specialization also attracted other sources of agglomeration economies not often found in rural areas, such as learning by training apprentices through guilds (De la Croix et al., 2018) as well as the matching of those apprentices between guilds.[18]

Yet, with the aforementioned description about what city rights entail, a sudden transition from a subsistence farming, feudal society towards a manufacturing urban society due to city rights indicates a structural break point from which point onwards second-nature forces contributed relatively more and more. As such, and even though we cannot measure

[18] Guilds also set and ensured quality standards as well as provided basic social security.

the relative share of first- and second-nature contributions, we can nonetheless think of awarded city rights as the period where the impact of first-nature geography starts to decline relative to second-nature geography. We are not aware of an alternative demographic indicator that is systematically available for all cities at an earlier development stage (see also Barjamovic et al., 2019).

The second wave of city rights started during industrialization and were typically given if the place reached a certain size, or there was merely the wish of the place to be called 'city'. To abstract from the impending urbanization and industrialization, which has less to do with some inherent locational advantage, we will therefore only consider city rights that were issued before 1700 in **Table 5**, where we estimate the Zipf coefficient of our geographical advantage proxy.

The probability of the inverse early geographical advantage gap_i for some city i being smaller than some A is equal to

$$\Pr(gap_i < A) = \frac{c}{A^{\eta}} \tag{5}$$

For some positive constant c and η. We regress the following equivalent log-linear version through ordinary least squares (OLS):

$$\log\left(rank(gap_i) - \frac{1}{2}\right) = \alpha + \eta \log(gap_i) \tag{6}$$

and interpret a positive slope η being estimated close to 1 and with a straight line, i.e., high R^2, as the gap of city rights and first record year following a Zipf distribution. As before, we have adjusted the rank by -0.5 to correct for potential small sample bias (Gabaix and Ibragimov 2011). Gabaix and Ioannides (2004) showed with a simple model that, if η_{pop} is the power law exponent of the population distribution across, in our case, places, and we show $\eta_{pop} = \eta = 1$, then geographical advantages can be an explanation of Zipf's law for cities/places.[19,20]

<< TABLE 5 ABOUT HERE >>

[19] We may additionally require that this equality holds at a given time, so that we require $\eta_{pop,t} = \eta_t$, if $\eta_{pop} \neq 1$ at the time of interest t. We do not have this population data for this many places around the Middle Ages, but the literature suggests that Zipf's law is a stable empirical regularity throughout history. Hence, we consider an estimated $\eta = 1$ as not the definitive answer on whether geographical advantages explain Zipf's law, but a strong suggestion that it does.

[20] Barjamovic et al. (2019) inferred population sizes from ancient Assyrian trade data and a structural model, and found a Zipf distribution for the estimated population of some of the largest cities during the Bronze Age in Anatolia. Our approach does not directly measure actual population sizes as well; our proxy can be interpreted as the growth rate between a primitive settlement until a mature city.

Table 5 shows that our measure for geographical advantages of early settlements indeed exhibits a Zipf distribution. Furthermore, similar to the frequency of characteristics in section 4.1, we observe that the behavior of the coefficient approaches unity as we decrease the maximum difference. **Figure 5** shows the relation of the rank of the gap and the gap itself. Notice that for high values of the gap, i.e., low-geographical advantage places, the points do not fit the linear fit line perfectly. A simple explanation is as the book only considers places above around 7,500 population today, we would have excluded relatively more places with a low level of geographical advantage. Simply expanding the sample would likely create a better picture in the upper tail. An additional explanation for this deviation is that those places represent cities that were founded much earlier, before the first modern city rights around 1150 spread throughout the region. We can control for this by adding bicentennial dummies for the first recorded mention, starting from year 1AD in panel B of **Table 5**. The estimates are now even closer to unity as compared to the specification without controls of panel A. We will hence control for this in the remainder of the paper when appropriate.

<< FIGURE 5 ABOUT HERE >>

If we take our 'gap' proxy as a measure of early settlement geographical advantage, the results in **Table 5**, together with the results of section 4.1, support the conjecture of geographical advantages being itself Zipf distributed as proposed by Krugman (1996). These results may also open up alternative explanations for Zipf's law that do not rely on some form of Gibrat's law with a reflexive lower bound (Gabaix 1999b, Eeckhout 2004, Duranton 2006, Rossi-Hansberg and Wright 2007). For example, traditional urban system models (Henderson 1974) may be compatible with a Zipf distribution by assuming underlying determinants that are themselves Zipf distributed (Henderson, 1988; Gabaix and Ioannides, 2004), as we demonstrated with geographical characteristics in this section.

More recently, Lee and Li (2013) develop a static model where the city size is determined by a product of multiple random factors which are i.i.d., a subset of these factors being represented by heterogeneous exogeneous physical features. Importantly, they generate a Zipf distribution without the need for Gibrat's law as well. In any case, we have shown that the empirical regularities pertaining to city-size distributions also have a bearing on the first-nature geography of places or cities. In the next section, we will investigate which specific geographical characteristics help to explain the size of German settlements in modern times.

5. Geographical characteristics and local economic activity

5.1. Impact of early characteristics and modern city size

We now turn to the question as to whether and how the initial geographic characteristics explain variation in 1910 city sizes. We estimate the impact of the geographic characteristics suggested by a toponym, compared to those that are without any geographic meaning, on the 1910 population of place i in country c. That is, we estimate:

$$\log(\mathrm{pop_{ic}})_{1910} = \alpha + \mathrm{X_{ic}}\beta + \gamma_1 \mathrm{NoGeo_{ic}} + \gamma_2 \log(\mathrm{UP_{ic}}) + \gamma_3 \mathrm{FirstRecord_{ic}} + \mathrm{Z}\delta + \varepsilon_{\mathrm{ic}} \quad (7)$$

where X is a vector of the binary first-nature characteristics, $NoGeo$ is the binary variable aggregating places where the suffix of the toponym does not have a geographic meaning, and UP is our urban potential control on the 1910 population data, described in section 3.3. $FirstRecord$ is the year in which the place name was first recorded.[21] Z is a vector of control variables accounting for latitude, longitude, their squared and interaction terms, and a dummy whether a place had city rights before 1910. In all regressions, we include country fixed effects to account for fixed and unobservable country-level characteristics that may uniformly affect places within national borders.

Our treatment assignment, which in this case are geographical characteristics as implied by a place's toponym, may be regionally clustered. For example, the southern regions of Germany are relatively more mountainous, while the region of Friesland could be considered more swamp- or bog-like. Hence, the treatment assignment may be clustered. Following Abadie et al. (2017), however, the appropriate level to cluster in this case is at the individual place level and not, for instance, at a district level. Note that due to our cross-sectional data structure, individual level clustered standard errors are the same as heteroskedasticity robust standard errors.

Table 6 presents our baseline estimates, where we show the parameter estimates for 33 of the most common geographical characteristics in our sample of 3,705 places. In columns (1)-(6), we introduce our first-nature characteristics in alternating order of the overarching categories: Water, Structures, Vegetation, Land, Elevation, and Resources.[22] Note that bridges may also fit in the Water category instead of Structures, while hedges may not have developed naturally, but were often man-made structures as a primitive form of protection. A similar argument can be made for floodplains belonging to the land category, as well as clearings belonging to the resources and/or land category. Thus, the purpose of these

[21] Following Giesen and Suedekum (2014), U.S. settlements that were founded earlier are likely larger in present times. Hence, some of the coefficients may be biased upwards by earlier settlements being founded disproportionally around some characteristics. Therefore, we add the first recorded mention in column VII to control for this. To a certain degree, this will also control for some regional differences, as settlements were initially in the regions west of the Rhine, with a wave of settlements towards the North East only happening after around 800AD with the Frankish Colonization and later the *Ostsiedlung* (Eastern Settlement).

[22] All regressions control for toponyms in the Miscellaneous category, of which the underlying characteristics are relatively uncommon and their parameter estimates omitted from the tables.

overarching categories is primarily for the convenience of the reader. The main results without this somewhat arbitrary categorization are in column (7), where we consider all first-nature characteristics in our sample.

Note that compared to places with geographic characteristics in their toponyms, places without geographic meaning have ($e^{-0.015} - 1 \approx$) 1.5% lower 1910 population levels. As expected, urban potential correlates positively with urban development. Secondly, places with city rights awarded prior to 1910 are ($e^{0.130} - 1 \approx$) 14% larger in 1910 compared to places without city rights, all else equal. Moreover, the sign for the first-record year is negative and consistent with the findings for the U.S., where city age is positively correlated with city population size (Giesen and Suedekum, 2014).

According to the results in **Table 6**, there is considerable heterogeneity in the extent to which various characteristics within each of the categories correlate with a place's size. However, the point estimates across the columns are relatively stable. Based on our preferred regression in column (7), we find that toponym references to rivers and fords correlate positively with 1910 population levels. The results indicate that places in 1910 that refer to rivers are on average larger than places referring to streams. Furthermore, places with a ford that may serve natural trading hubs through land and water show an even larger effect. That is, as compared to the baseline case of no geography, the early presence of fords indicates a size advantage in 1910 of about 8.3 percent. This indicates the importance of characteristics associated with trade as compared to those merely associated with having access to water.

As for structures, we find a lasting size advantage of the presence of early churches of about 7.8 percent as compared to the no geography baseline, but an insignificant effect for castles and other man-made structures. Especially the insignificance of bridges may be surprising given its similar role with fords. We rationalize this by noting that the typical bridge of early settlements during the Middle Ages were likely not of the sophisticated Roman stone bridge type, and were more likely to have disappeared or become out of use by 1910, unlike fords. In the Vegetation category we find a negative relation of early hedges with 1910 population size, but a lasting size advantage for places associated with a clearing. We should note here that clearings may also indicate locations where forests were cleared for wood and agricultural purposes. An advantage of this land is that the roots remaining in the soil prevent erosion and may provide nutrient rich soil in case of 'slash-and-burn' type of clearings.

In terms of elevation, we find that hills have a positive and significant impact on modern (= 1910) city population size. We rationalize this finding in the sense that hills provide an elevated vantage point for defensive purposes and also provide safety from local flooding. Finally, none of the characteristics categorized in Land or Resources are significantly related to 1910 population levels. Again, we should note that clearings could also be considered as belonging to the Land category, while bogs provide peat as a resource for heating.

<< TABLE 6 ABOUT HERE >>

Conceptually, one may be concerned that both our measure of natural geographical advantages, proxied by place names, and urban growth are likely to be correlated with a place's underlying trade potential. That is, are cities more populous due to their trade-facilitating first-nature characteristics or because of the man-made trade potential derived from these first-nature advantages? Ideally, one would need to control for a place's time-varying trade potential to rule out this possible threat to identification.

In the absence of reliable historical trade data for the places in our dataset, we construct a time-varying proxy measure that controls for its access to other nearby markets, MA. Our measure of trade potential relates to Bakker et al.'s (2021) measure of connectedness, and Bosker and Buringh's (2017) measure for second-nature geography. In our case, we exploit the fact that there *cannot* be any legal trade in places that do not have city rights.

Specifically, for each place, we calculate the number of other places with city rights obtained within a certain period of time and within a given geographic radius. We construct three different distance bands, measuring the number of cities with city rights (i.e., markets) within a distance of 20km, 20-50km, and 50-100km. These distance bands correspond, respectively, to a 1 day, 2.5 day, and 5-day return trip in our sample period (Bosker and Buringh, 2017). These measures are time-varying in that we restrict our calculations only to those places that obtained city rights before 1200, before 1300, and so on.

Similar to the urban growth shadows literature (Bosker and Buringh, 2017; Cuberes et al., 2021), we expect that for places that are geographically close to many other markets (i.e., places with city rights), the economic viability or necessity of establishing yet another market is more limited compared to those places that are more distant from other markets, because for those markets the benefits are larger. Thus, our measure of a place's literal access to other markets reflects its trade potential.

We then re-estimate equation (7) and include our period-specific distance bands that control for the (log) number of places with city rights.[23] The results are presented in **Appendix Table B1**. For ease of comparison, column (1) repeats our baseline estimates from **Table 6**, column (7). Column (2) adds the trade potential proxy based on places that obtained city rights before 1200, column (3) for places with city rights before 1300, and so on.

Looking across the results in **Appendix Table B1**, two main observations stand out. First, the proxies for trade potential are all statistically significant and with the expected negative sign: urban growth in terms of log 1910 population is lower for places in close proximity to other markets. Second, and crucially: the point estimates for our geographical characteristics are very similar to our baseline results. That is, controlling for a place's inherent trade potential does change our key finding that urban growth can be explained by the geographical information contained in our toponymical data.

Next, we turn to potential concerns regarding the underlying toponymical data, which we address in a series of sensitivity checks of which the results are reported in **Table 7**.

[23] Following Raballand (2003), we take log(MA + 0.01) to avoid dropping observations where the market access proxy is 0.

<< TABLE 7 ABOUT HERE >>

The first concern is that the crucial suffix part of a place name (see section 3.2) may be subject to different interpretations. As for the coding into our dataset, we rely on the preferred, likely or consensus interpretation as indicated in the book, and dismiss the fringe interpretations. Yet, there still might be some measurement bias in the meaning of these toponyms.[24] All toponyms with multiple interpretations in that manner are therefore excluded in column (1).

Second, and as a follow-up to column (1), some toponyms may be subject to multiple interpretations where no consensus, clearly preferred or any interpretation at all has emerged. In that case, we do not assign any characteristic to that place and all toponyms without a clear interpretation are excluded in column (2).

Third, we may worry about the toponyms that contain both a geographic (typically prefix) as well as a non-geographic (typically suffix) element, instead of the usual constellation of a prefix that further describes the geographic suffix itself. These places are excluded in column (3).

Fourth, there could be measurement bias in the first-record year, particularly for cases when the book gave imprecise dates such as "around 900AD" which we interpreted as "900AD", and "15th century" as 1450. Column (4) drops all cases in which the precise first-record years do not directly follow from the book or other archival records.

Finally, our most restrictive specification in column (5) is one in which we drop all observations subject to at least one of four above-mentioned limitations.

Overall, most of our earlier conclusions remain robust to these various sensitivity checks. In particular, we find a robust and sizeable, positive impact of rivers, fords, churches, fences, clearings, and hills on 1910 population levels. We support the prevalent notion within the literature (see **Table 1**) of the importance of water-based trade and trade hubs in particular. However, to say something about the persistence of these effects and path dependence, we need to estimate their importance also early in a settlement's life, which is the topic of the next sub-section.

5.2. Impact of early characteristics on early geographical advantage

We will repeat the analysis of section 5.1, but now with the dependent variable being our measure of early settlement geographical advantage as introduced in section 4.2. The idea is to uncover the determinants of early geographical advantages as proxied by our 'gap' variable, see equation (4). That is, we estimate:

$$\log(\mathrm{gap}_{ic}) = \alpha + \mathrm{X}_{ic}\beta + \gamma_1 \mathrm{NoGeo}_{ic} + \mathrm{Z}\delta + \varepsilon_{ic} \qquad (8)$$

[24] We suspect that largest cities are more often subject to multiple interpretations due to the positive correlation of city size today and city age, where older toponyms are more often less clear in the interpretation. Furthermore, we suspect that there is a higher interest in the interpretation of larger cities, which invites fringe theories on the interpretation of the toponym especially of larger cities. A simple, bivariate regression of log 1910 population and the multiple interpretation indicator yields a highly significant coefficient of 0.336 with a 0.098 standard error.

where Z now is a vector of control variables for latitude, longitude, their square and interaction terms, and a bicentennial dummy. As before, these regressions are estimated with (current border) country fixed effects. In this case this is done to primarily control for the contributors of the book being potentially biased towards interpreting toponyms of places within the current borders of Germany.

Table 8, column (1) shows the baseline results. The remaining columns repeat the robustness checks that were introduced in section 5.1 to control for potential biases that might arise from suffixes with multiple possible interpretations (column 2), without any clearly preferred, consensus or any interpretation (column 3), toponyms with mixed geographical and non-geographical meanings (column 4), places with imprecise first-record years (column 4), or a combination of all these concerns (column 5).

In column (1), the negative coefficient for castles suggests a smaller gap between the first-record and year when city rights were awarded, all else equal. In other words, places near castles are awarded city rights sooner on average after their first mention, all else equal. The estimated coefficients suggest a significant early geographical advantage of castles, bridges and, once omitting observations as in columns (2-5), also of lakes. Nevertheless, as per **Table 7**, the early geographic advantage of castles, bridges and lakes for cities does not have a lasting impact on modern-day city growth. Moreover, we find that the pattern has reversed for fords and bridges, which we noted to be similar in function, with the results suggesting that fords yielded less of an early geographic advantage compared to bridges.

On the other hand, the early geographical disadvantage of being located on a river island and being near to a (spring) source, church or a farm estate also did not last until 1910. However, we should also note that we estimate a positive and significant coefficient for the no-geography indicator, so that the estimated disadvantages are smaller compared to the no-geography baseline. We also note that castles and river islands both offer security, but the estimated coefficients are in stark contrast to another. As city rights allow for markets, the larger average 'gap' may simply reflect river islands to be difficult to access and to sustain a market on, offsetting the effect of added security.

<< TABLE 8 ABOUT HERE >>

As in the previous section, we address the possible concern of trade-related endogeneity in our baseline estimates by running additional regressions that include our time-varying proxy for a place's market access (MA) to other places with city rights at various distances. These results are presented in **Appendix Table B2**. Comfortingly, the point estimates for our characteristics are quantitatively and qualitatively very similar to the baseline estimates. Moreover, these results suggest that higher competition through more nearby markets (in the 0-20km distance band) lowers their early geographical advantage, all else equal. In line with the findings on urban chances by Bosker and Buringh (2017), we also estimate that a higher number of markets farther away at the 50-100km distance band is associated with a

shorter gap between the year when city rights were awarded and the first-record year, which indicates that the viability/necessity for trade and not increasing competition dominates at this distance band.

Overall, these findings and those of the previous subsection indicate a shift in the importance of land-based trade through bridges, notably towards water-based trade through rivers over time. These results align with the literature that stresses trade as an important determinant of early settlement development, such as Bakker et al. (2021) for the ancient Phoenician trade in the Mediterranean Sea, as well as the change in the primary mode of transportation through technology (Bosker et al., 2013; Michaels and Rauch, 2018). Our results complement the results found in that literature with a few key distinctions: our dataset mostly consists of settlements that developed during the Middle Ages and include many smaller places that still exist today, so that we can evaluate the continuous settlement history. The settlements in our sample also lack good coastal access as compared to the Mediterranean Ancient World, so that trade within the Holy Roman Empire likely took place relatively more through land routes as compared to sea-routes.

5.3. Heterogeneous reaction of characteristics with agglomeration and modern population

From **Figure 2**, we know that most city rights were awarded in the first wave before the 16th century, with a second wave of city rights being awarded from the 19th century onwards. One could sum up these periods as the extensive margin of population being the driver in the first wave with new cities arising, for example with the Frankish Colonization and later the *Ostsiedlung* (Eastern Settlement) as two concrete examples. On the other hand, the intensive margin could be seen as the driver during the second wave, as there were few new settlements but existing cities grew. Note that a new wave of city rights was awarded from the start of industrialization in the 19th century, further increasing urbanization.

Similar to Henderson et al. (2017), we can think of early, first-wave and late second-wave cities in this case, and examine their differential effect on city growth depending on the characteristic. Older cities are likely to be larger today, for example as shown by the negative first-record control coefficients in **Table 6**, so that we are more likely to find a positive differential effect of early cities. However, this may depend on the characteristic itself, with differences here telling us how different characteristics make use of population, and thus second-nature geography over time. We estimate the following:

$$\log(pop_{ic})_{1910} = \alpha + \boldsymbol{X}_{ic}\beta + Early_{ic}\,\boldsymbol{X}_{ic}\,\beta_d + \gamma_1 NoGeo_{ic} + \gamma_{1d}\,NoGeo_{ic}\,Early_{ic} + \gamma_2 log(UP_{ic}) + \gamma_3 FirstRecord_{ic} + \boldsymbol{Z}\delta + \varepsilon_{ic} \quad (9)$$

which is equivalent to equation (7), except that we now include the interaction terms, with $Early_{ic}$ being a dummy variable indicating whether a place received city rights before 1800, and which we call an 'early' city. Note that a 'late' city in this definition does not only consider those that received its city rights on and after 1800, but also those that did not receive city rights so that we again consider the full sample as in subsection 5.1.

<< TABLE 9 ABOUT HERE >>

We consider 1800 specifically as being right before the wave of industrialization and urbanization. In general, if the interaction effect exceeds the main effect, it suggests that there is a magnifying effect of agglomeration on this characteristic, which we report for most variables. In **Table 9**, column (1) reports the main baseline effect for late cities and column (2) the interaction effects which shows the differential for early cities, obtained from the same regression. The results reported in columns (3) and (4) incorporate the robustness checks we introduced in section 5.1., i.e., to control for potential biases that might arise from a suffix with multiple possible interpretations, without a preferred interpretation, names with combined geographical and non-geographical meanings, and places with imprecise first-record years. For both specifications, we note that the ratio of main and interaction of the non-geographic meaning is about -0.59. We attribute this to agglomeration effects, since it is driven by underlying effects not related to some early natural advantage, at least for the case of places without first-nature geography characteristics in the toponym.

We see the positive effect of floodplains, sources, streams and rivers on urban growth only applying to early cities urbanized before industrialization, but not for late cities. It seems that the benefits of these characteristics are positively coupled to the agglomeration of a location. Taking the example of rivers, we may explain this by the vast number of places along rivers and that not every place, no matter its size, makes equal use of rivers to become a trading hub due to local saturation. As for fords, we see the most pronounced effect for late cities with a positive but insignificant effect for early cities. That is, fords as a hub thrived despite not being early agglomerated locations, in contrast to rivers. And unlike rivers, fords are a rare element of first-nature geography in our setting, which may not have been affected by local saturation.

Similarly, we find heterogeneity with fields, bog, and hedges that are positive for early cities in contrast to the negative estimates for late cities. The purpose of hedges was not only for some form of initial protection, they were also as a funnel for toll collection (Küntzel, 2004) and therefore important for trade routes. Bogs, on the other hand, like swamps, are challenging to build on and were only used for settlement much later.[25] However, unlike swamps, bogs do provide peat as a fossil fuel primarily used for heating, which was a reason settlement started there during the late Middle Ages. One explanation for the large heterogeneity for bogs could be that early bog cities were founded specifically for the extraction of peat, unlike late bog cities.

The case for places that developed out of initial clearing activities is also interesting since clearing activities as a form of historical accident do not offer a permanent natural advantage, so that we expect modern patterns to be driven by path dependence. We also do not find that such activities constitute an early advantage in the previous subsection. However, we do

[25] While the average first recorded mention for our sample is in 1098 and for city rights in 1520, these events happened much later for bog places at around 1508 and 1696, respectively.

find a positive effect of this historical characteristic for early cities, suggesting strong heterogeneity driven by whether the past clearing activities were sufficient to propel the location to becoming an early city.

Finally, we find that early cities around mountains and elevations have a pronounced positive effect on 1910 population levels, but this effect disappears for late cities. Taken together, our results indicate a large heterogeneity of reaction between and within various categories of first-nature characteristics. In particular, we show that proximity to streams, sources, floodplains, rivers, clearings, hedges, elevation, fields and bogs propelled early cities towards urban agglomeration.

6. Conclusion

Most cities in the world can trace their origins back to some modest beginnings. Initial geographical characteristics were often important for city development. We are in particular interested in the importance of initial conditions surrounding a settlement around the time of its initial development. Importantly, with human and natural interventions, the geographical circumstance at the initial time of settlement might have changed over time.

Place names (or so-called toponyms) provide a unique glimpse into the past, as they preserve in their name the original geographic surroundings of the early settlement over many centuries, which is an advantage over measures using present-day geography.

Our main research question is how early geographical characteristics contribute to subsequent city growth, and how the importance of these characteristics changes over time. To this end, we constructed a unique dataset with the interpretation of the toponyms of 3,705 (mostly) German settlements, revealing its immediate early surroundings.

One of our main contributions to the literature is that we vastly expand on the number and type of first-nature geography characteristics, with a total of 168 characteristics. In addition, we digitized the first recorded mention of a settlement, as well as the year in which a settlement received city rights. We use these data to construct a simple measure for early settlement geographical advantage, where the rationale is that settlements with better early geographical advantages would receive city rights faster after being first mentioned in any record.

Our main findings are as follows. Thanks to the expanded number of different first-nature geography features, we find that the frequency of these geographical characteristics follows a Zipf distribution. We also find that our early settlement geographical advantage measure follows a Zipf distribution as well. While the former result is interesting by itself, the latter is important to the literature on the city size distribution, as it is the first evidence for a Zipf's law of local geographical advantage that has long been suggested as an alternative explanation for Zipf's law for cities (Krugman 1996, Davis and Weinstein 2002, Gabaix and Ioannides 2004). Second, we estimated the impact of the implied early geographical characteristics of a settlement on its population in 1910. Furthermore, using the proxy measure for early geographical advantage, we find that the early settlements next to castles

and lakes obtained city rights relatively quickly, in stark contrast to the early settlements next to streams, sources, river islands, churches or farm estates.

However, we also show that some early geographical characteristics, particularly the access to some waterway (rivers, fords) became a more important feature in modern days. Similarly, places that had an early presence of churches as well as an early historical activity of clearing of forests are on average larger today. Overall, these results are consistent with the literature that emphasizes the role of water-based trade characteristics and its rising importance over time (e.g. Bosker et al. 2008, Bosker et al. 2013, Henderson et al. 2017, Bakker et al. 2019, Düben and Krause 2023). Lastly, we do find a heterogenous relationship between some of the geographical characteristics and agglomeration. We find that the population of places along rivers, floodplains and sources benefitted from being an early agglomeration, while we find the opposite case for fords or river mouths. Another distinction to previous studies is that our dataset allows us to not only focus on the largest Medieval cities (cf. Bairoch et al 1988), but many other smaller settlements, particularly during their infant development stage where we typically think of historical accidents to be the most influential in steering the course of a settlement's development. In line with Düben and Krause (2023) in their historical analysis of the relevance of physical geography for Chinese urban growth, we show how initial physical geography, as proxied by toponyms in our case, might matter across a very wide range of city sizes.

We see a future avenue of research in the estimation of how individual geographical characteristics change in importance over time. This would remedy a limitation of the present study, but it requires access to better population estimates throughout time; better in terms of how far the population data can go back in time, as well as how many places are considered (see also Chaney, 2020). In a similar vein, it would be very interesting to see whether a toponym-based approach can be used to unravel the relevance of initial geography for urban development in other historical settings other than the German context.

Acknowledgements

We thank Maarten Bosker, Jacopo Canello, Donald Davis, Yannis Ioannidis, Jeffrey Lin, Dimitrios Soudis, Romain Wacziarg, and participants of the 2020 Meeting of the Urban Economics Association and 2021 Summer Conference in Regional Science of the German-speaking section of the European Regional Science Association for helpful comments.

Declarations of interest

None.

Data and code disclosure

A complete replication package containing the complete dataset and Stata .do files used to generate the results reported in this study will be publicly available on the DataverseNL (dataverse.nl) data repository.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (Working Paper No. 24003). National Bureau of Economic Research.

Acemoglu, D., Johnson, S., and Robinson, J. (2005). The rise of Europe: Atlantic trade, institutional change, and economic growth. *American Economic Review*, *95*(3), 546–579.

Allen, T., and Donaldson, D. (2020). *Persistence and path dependence in the spatial economy* (Working Paper No. 28059). National Bureau of Economic Research.

Alsan, M. (2015). The effect of the tsetse fly on African development. *American Economic Review*, *105*(1), 382–410.

Armenter, R., Koren, M., and Nagy, D. K. (2014). Bridges. *Unpublished Manuscript*.

Bairoch, P., Batou, J., and Chevre, P. (1988). La population des villes Europeennes, 800-1850. *Librairie Droz, Geneve*.

Bakker, J., Maurer, S., Pischke, J., and Rauch, F. (2021). Of mice and merchants: connectedness and the location of economic activity in the Iron Age. *Review of Economics and Statistics*.

Barjamovic, G., Chaney, T., Cosar, K., and Hortacsu, A. (2019). Trade, merchants, and the lost cities of the Bronze Age. *The Quarterly Journal of Economics*, *134*(3), 1455–1503.

Berliant, M., and Konishi, H. (2000). The endogenous formation of a city: population agglomeration and marketplaces in a location-specific production economy. *Regional Science and Urban Economics*, *30*(3), 289–324.

Bleakley, H., and Lin, J. (2012). Portage and path dependence. *The Quarterly Journal of Economics*, *127*(2), 587–644.

Bosker, M., Brakman, S., Garretsen, H., De Jong, H., and Schramm, M. (2008). Ports, plagues and politics: explaining Italian city growth 1300–1861. *European Review of Economic History*, *12*(1), 97–131.

Bosker, M., and Buringh, E. (2017). City seeds: Geography and the origins of the European city system. *Journal of Urban Economics*, *98*, 139–157.

Bosker, M., Buringh, E., and Van Zanden, J. L. (2013). From Baghdad to London: Unraveling urban development in Europe, the Middle East, and North Africa, 800–1800. *Review of Economics and Statistics*, *95*(4), 1418–1437.

Brakman, S., Garretsen, H., Van Marrewijk, C., and Van Den Berg, M. (1999). The return of Zipf: Towards a further understanding of the rank-size distribution. *Journal of Regional Science*, *39*(1), 183–213.

Brendler, S. (2008). Sprachliche Orientierung im Raum: Zur Notwendigkeit der Ortsnamen. *Linguence Hamburg*.

Chaney, E. (2013). Revolt on the nile: Economic shocks, religion, and political power. *Econometrica*, *81*(5), 2033–2053.

Chaney, E. (2020). Modern library holdings and historic city growth. *CEPR Discussion Paper No. DP14686*.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration economics* (pp. 15–66). University of Chicago Press. Cronon, W. (1991). *Nature's metropolis: Chicago and the Great West*. New York: W. W. Norton.

Cuberes, D., Desmet, K. and Rappaport, J. (2021). Urban Growth Shadows. *Journal of Urban Economics,* 123, 103334.

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.

Dalton, J., and Leung, T. (2014) Why is Polygyny More Prevalent in Western Africa? An African Slave Trade Perspective. *Economic Development and Cultural Change*, 62(4): 599–632.

De la Croix, D., Doepke, M., and Mokyr, J. (2018). Clans, guilds, and markets: Apprenticeship institutions and growth in the preindustrial economy. *The Quarterly Journal of Economics*, *133*(1), 1–70.

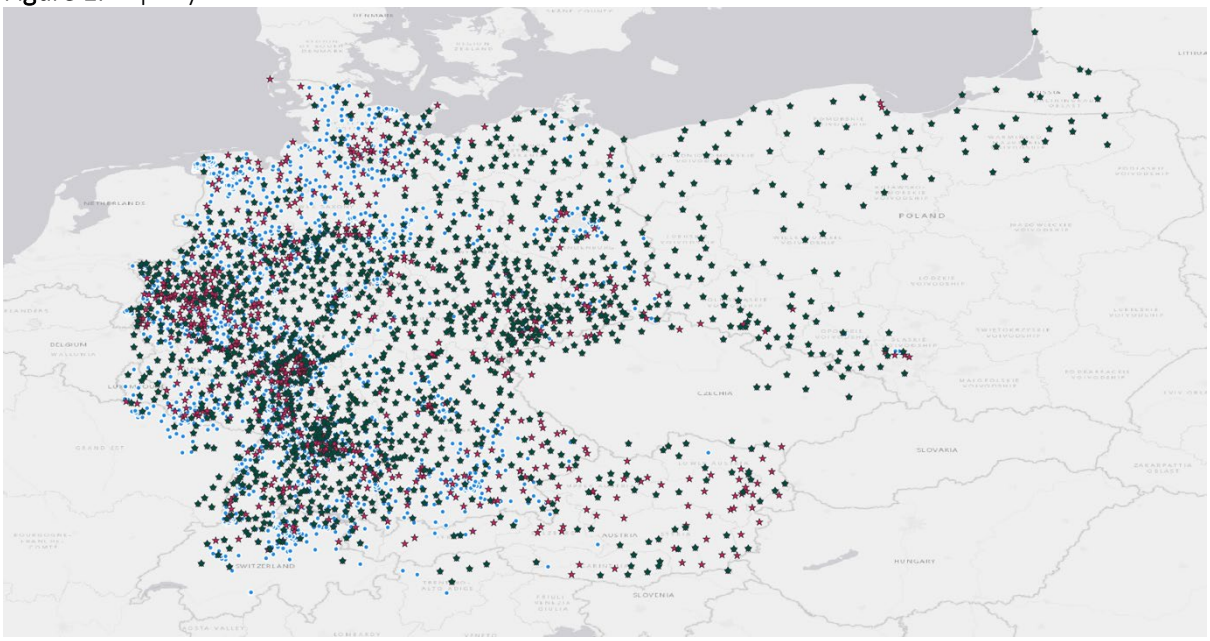
Dittmar, J. (2011). Cities, markets, and growth: the emergence of zipf's law. *Unpublished Manuscript*.

Düben, C., & Krause, M. (2023). The Emperor's Geography—City Locations, Nature and Institutional Optimisation. *The Economic Journal*, 133(651), 1067-1105.

Duranton, G. (2006). Some foundations for zipf's law: Product proliferation and local spillovers. *Regional science and urban Economics*, *36*(4), 542–563.
Duranton, G., and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063–2117). Elsevier.
Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, *94*(5), 1429–1451.
Fenske, J., & Kala, N. (2015). Climate and the slave trade. *Journal of Development Economics*, *112*, 19-32.
Flückiger, M., Hornung, E., Larch, M., Ludwig, M., and Mees, A. (2019). Roman transport network connectivity and economic integration. *CEPR Discussion Paper No. DP13838*.
Gabaix, X. (1999a). Zipf's law and the growth of cities. *American Economic Review*, *89*(2), 129–132.
Gabaix, X. (1999b). Zipf's law for cities: an explanation. *The Quarterly Journal of Economics*, *114*(3), 739–767.
Gabaix, X., and Ibragimov, R. (2011). Rank- 1/2: a simple way to improve the ols estimation of tail exponents. *Journal of Business and Economic Statistics*, *29*(1), 24–39.
Gabaix, X., and Ioannides, Y. M. (2004). The evolution of city size distributions. In *Handbook of regional and urban economics* (Vol. 4, pp. 2341–2378). Elsevier.
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.
Giesen, K., and Südekum, J. (2014). City age and city size. *European Economic Review*, *71*, 193–208.
Hanlon, W. W. (2020). Coal smoke, city growth, and the costs of the industrial revolution. *The Economic Journal*, 130(626), 462-488.
Hanlon, W. W., and Heblich, S. (2020). History and Urban Economics. *National Bureau of Economic Research Working Paper Series* (w27850).
Henderson, J. (1974). The sizes and types of cities. *The American Economic Review*, *64*(4), 640–656.
Henderson, J. (1988). Urban development: Theory, fact, and illusion. *Oxford: Oxford University Press*.
Henderson, J., Squires, T., Storeygard, A., and Weil, D. (2017). The global distribution of economic activity: Nature, history, and the role of trade. *The Quarterly Journal of Economics*, *133*(1), 357–406.
Hornbeck, R. (2012). The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe. *American Economic Review*, *102*(4), 1477–1507.
Ioannides, Y. M., and Overman, H. G. (2003). Zipf's law for cities: an empirical examination. *Regional Science and Urban Economics*, *33*(2), 127–137.
Ioannides, Y. M., and Overman, H. G. (2004). Spatial evolution of the US urban system. *Journal of Economic Geography*, *4*(2), 131–156.
Irsigler, F. (2015). Libertes, privileges et autogestion des populations urbaines. quelques conclusions d'un colloque d'histoire urbaine. In *Urban liberties and citizenship from the Middle Ages up to now* (p. 263ff.). Retrieved from http://opac.regesta-imperii.de/id/2084857
Izdebski, A., Słoczyński, T., Bonnier, A., Koloch, G., & Kouli, K. (2020). Landscape change and trade in ancient Greece: evidence from pollen data. *The Economic Journal*, 130(632), 2596-2618.
Jedwab, R., Johnson, N. D., and Koyama, M. (2020). Medieval cities through the lens of urban economics. *Regional Science and Urban Economics*, 103598.
Ketterer, T., and Rodrígues-Pose, A. (2018). Institutions vs. 'first-nature'geography: What drives economic growth in Europe's regions? *Papers in Regional Science*, 97(S1): S25-S62.
Konishi, H. (2000). Formation of hub cities: Transportation cost advantage and population agglomeration. *Journal of Urban Economics*, *48*(1), 1–28.
Krugman, P. (1996). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International economies*, *10*(4), 399–418.
Küntzel, T. (2004). Stadt und Grenze-die Landwehr der Stadtwüstung Nienover im südniedersächsischen Kontext. *Archaeologia historica.*, *29*(1), 167–191.
Lee, S., and Li, Q. (2013). Uneven landscapes and city size distributions. *Journal of Urban Economics*, *78*, 19–29.
Lee, S., and Lin, J. (2018). Natural amenities, neighbourhood dynamics, and persistence in the spatial distribution of income. *The Review of Economic Studies*, *85*(1), 663–694.
Michaels, G., and Rauch, F. (2018). Resetting the urban network: 117–2012. *The Economic Journal*, *128*(608), 378–412.
Nagy, D. K. (2021). Quantitative economic geography meets history: Questions, answers and challenges. *Regional Science and Urban Economics*, 103675.
Niemeyer, M. (2012). Deutsches Ortsnamenbuch. *De Gruyter*.
Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics*, *57*(1), 86–100.
Nunn, N. (2008). The long-term effects of Africa's slave trades. *The Quarterly Journal of Economics*, *123*(1), 139-176.
Nunn, N., and Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, *94*(1), 20–36.
Oto-Peralias, D. (2017). What do street names tell us? The 'city-text' as socio-cultural data. *Journal of Economic Geography*, *18*(1), 187–211.
Rabbaland, G. (2003). Determinants of the negative impact of being landlocked on trade: An empirical investigation through the Central Asian case. *Comparative Economic Studies*, *45*(5), 520–536.
Rappaport, J., and Sachs, J. D. (2003). The United States as a coastal nation. *Journal of Economic Growth*, *8*(1), 5–46.
Redding, S. J. (2010). The empirics of new economic geography. *Journal of Regional Science*, *50*(1), 297–311.

Rossi-Hansberg, E., and Wright, M. L. (2007). Urban structure and growth. *The Review of Economic Studies*, *74*(2), 597–624.
Schilling, H. (2004). *Die Stadt in der Frühen Neuzeit*. De Gruyter Oldenbourg.
Soo, K. T. (2005). Zipf's law for cities: a cross-country investigation. *Regional Science and Urban Economics*, *35*(3), 239–263.
Stoob, H. (1985). *Die Stadt: Gestalt und Wandel bis zum industriellen Zeitalter.* Böhlau Verlag, Köln.
Tuckermann, W. (1906). *Das Gewerbe der Stadt Hildesheim bis zur Mitte des fünfzehnten Jahrhunderts*. E.Ebering Berlin.
Villette, J., and Purves, R. (2018). Exploring microtoponyms through linguistic and geographic perspectives. *21th AGILE Conference on Geographic Information Science, Lund*.
Villette, J., and Purves, R. S. (2019). From Microtoponyms to Landscape Using Semantics, Location, and Topography: The Case of Wald, Holz, Riet, and Moos in St. Gallen, Switzerland. *The Professional Geographer*, 1–12.

Figure 1. Toponym dataset



Note: Geographic coverage of the 3,705 toponyms in our dataset. Red stars indicate places that received city rights before 1800, green squares places that received city rights from 1800 onwards, while blue dots indicate places without city rights.

Figure 2. Distribution of first-record year and year when city rights were awarded

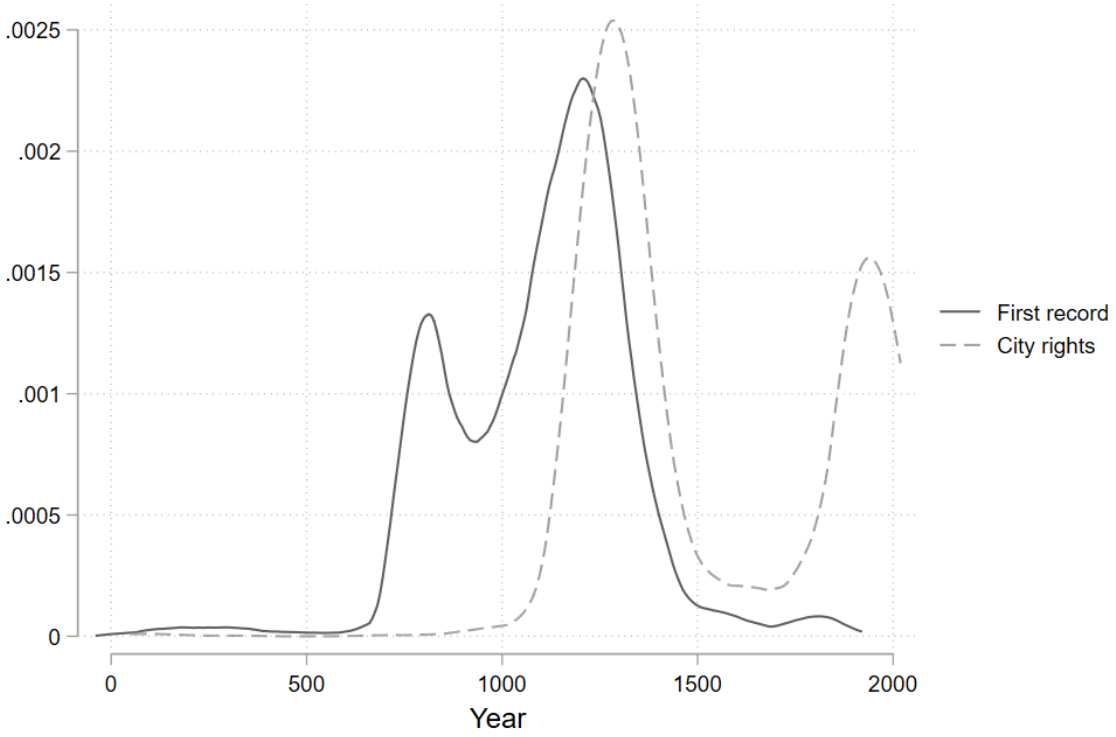


Figure 3. First-record distribution of toponymical characteristics

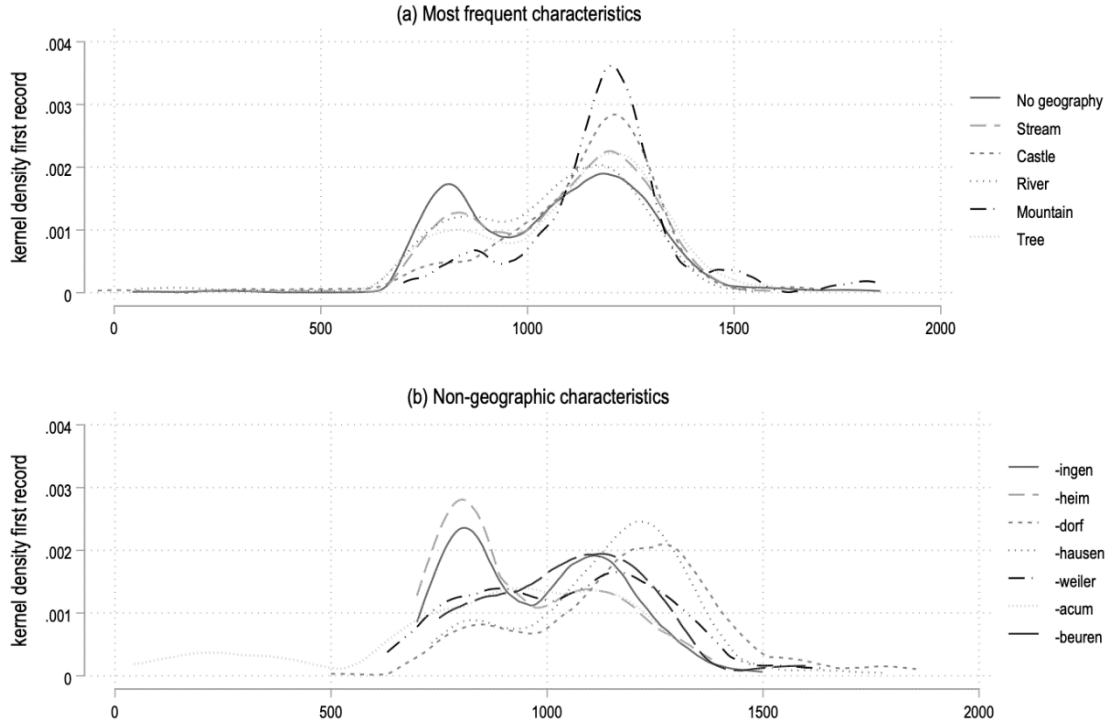
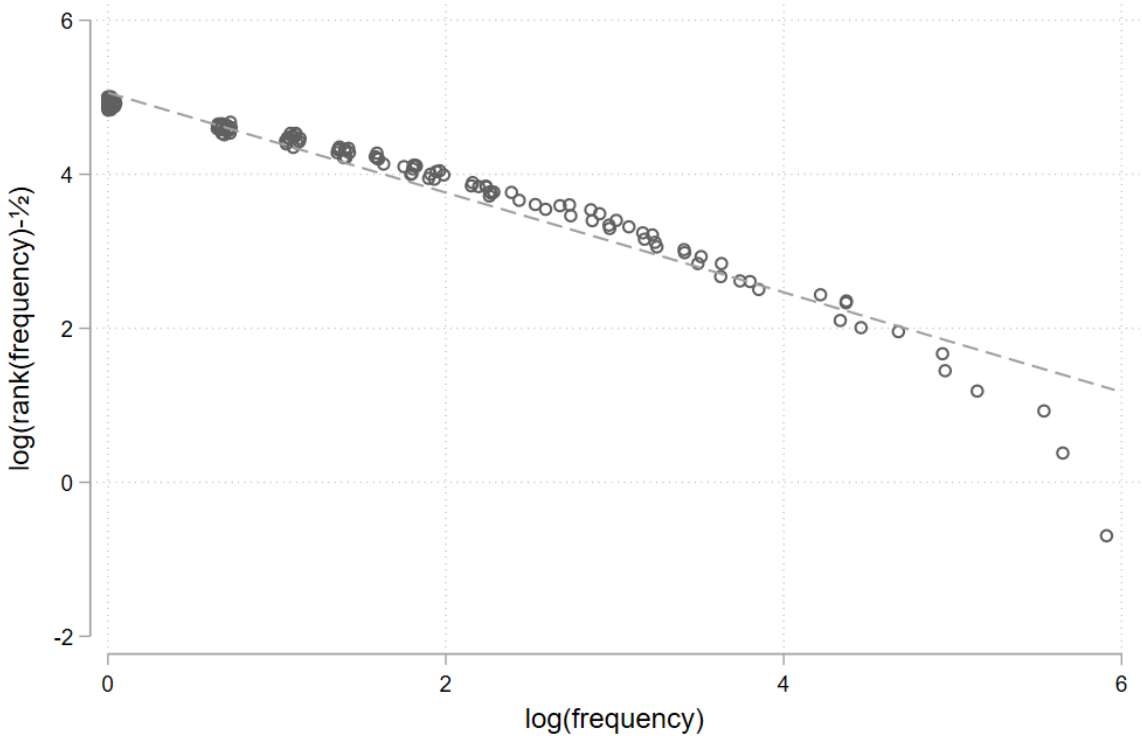
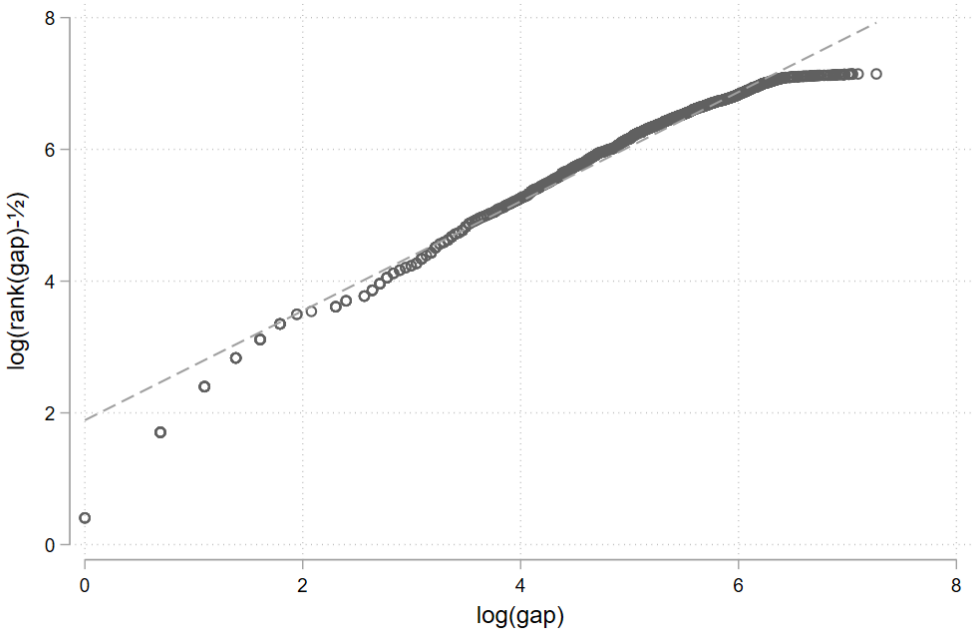


Figure 4. Zipf frequency of geographical characteristics



Note: Based on all places with geographic characteristics as per Table 4, panel A, without additional restrictions. We have added random noise (in STATA, option jitter(2)) to make overlapping points a little more visible.

Figure 5. Zipf gap for places receiving city rights before 1700



Note: Based on all places and geographic characteristics as per Table 4, panel A, without additional restrictions.

Table 1. Local economic activity and geography: Related studies

Study	Region	Period	Natural characteristics	Man-made characteristics	Main findings
Rappaport and Sachs (2003)	USA	2000	ocean coast, Great Lakes, navigable rivers		Coast, Great Lakes and navigable rivers have a positive effect on population density and employment density.
Acemoglu et al. (2005)	Europe	1300-1850		Atlantic ports	Effect of Atlantic ports after 1600 on city population points towards the importance of Atlantic trade.
Bosker et al. (2008)	Italy	1300-1861	Navigable waterways, mountains		Seaports and navigable waterways have a positive effect on urban development, mountains have a negative effect.
Bleakley and Lin (2012)	USA	1790-2000	Portage sites		Portage site places are still population centers today, even though portaging became irrelevant.
Bosker et al. (2013)	Latin West and Middle-East	800-1800	Seas, rivers		Positive effect of sea-access and Roman road hub in Latin West city population; negative effect of sea on Middle-Eastern cities. No effect of rivers.
Armenter et al. (2014)	USA	19th century			Post-bridge construction, places with bridges developed more than no-bridge places.
Bosker and Buringh (2017)	Europe	800-1800	Seas, rivers, ruggedness, elevation, cultivation probability		Water-based transportation increases in importance over time.
Henderson et al. (2017)	World	2010	Various biomes, temperature, malaria, precipitation, land suitability, elevation, river, coast, lake, ruggedness	Harbors	Harbors, rivers, lakes: positive effect on light intensity; negative effect for distance.

Table 2. Main categories and characteristics

WATER			STRUCTURES			VEGETATION			LAND		
Characteristic	*N*	*%*	*Characteristic*	*N*	*%*	*Characteristic*	*N*	*%*	*Characteristic*	*N*	*%*
Stream	387	41.0%	Castle	276	42.9%	Tree	147	27.5%	Field	108	32.2%
River	248	26.3%	Church	76	11.8%	Forest	137	25.6%	Swamp	44	13.1%
Floodplain	90	9.5%	Farm estate	67	10.4%	Clearing	79	14.8%	Bog	29	8.7%
Ford	38	4.0%	Monastery	45	7.0%	Hedge	76	14.2%	Land	27	8.1%
Source	37	3.9%	Bridge	25	3.9%	Bush	17	3.2%	Valley	21	6.3%
Lake	29	3.1%	Fence	24	3.7%	Heath	15	2.8%	Pasture	13	3.9%
Water	20	2.1%	Mill	20	3.1%	Forest free	13	2.4%	Wallow	11	3.3%
River mouth	17	1.8%	Roman castra	11	1.7%	Thicket	10	1.9%	Steep terrain	10	3.0%
River island	15	1.6%	Outlook tower	9	1.4%	Hurst	9	1.7%	Edge / Wedge	9	2.7%
Bath	7	0.7%	Walls	9	1.4%	Garden	7	1.3%	Bald spot	7	2.1%
Marsh	7	0.7%	Chapel	7	1.1%	Moss	5	0.9%	Mud	6	1.8%
Port	6	0.6%	Court	7	1.1%	Swamp forest	4	0.7%	Pile	6	1.8%
Weir	6	0.6%	Barn	6	0.9%	Tree stump	4	0.7%	Acre	5	1.5%
Boat	5	0.5%	Market	6	0.9%	Weathered plants	4	0.7%	Geest	5	1.5%
Canal	4	0.4%	Pole	6	0.9%	Twig	3	0.6%	Ravine	5	1.5%
Dam	3	0.3%	Road / Path	6	0.9%	Coppice	2	0.4%	Swampy meadow	4	1.2%
Fjoerd	3	0.3%	Hospital	4	0.6%	Unique cases	3	0.6%	Wasteland	4	1.2%
Hithe	3	0.3%	Warehouse	3	0.5%				Cover guard	3	0.9%
Island	3	0.3%	Drying place	2	0.3%				Fertile land	3	0.9%
Sluice	2	0.2%	Forestry	2	0.3%				Remote / Enclosed area	3	0.9%
Pool	2	0.2%	Iron works	2	0.3%				Trough	3	0.9%
Unique cases	11	1.2%	Shooting place	2	0.3%				Unique cases	9	2.7%
			Shop	2	0.3%						
			Smithery	2	0.3%						
			Tavern	2	0.3%						
			Tower	2	0.3%						
			Weavery	2	0.3%						
			Winery	2	0.3%						
			Unique cases	16	2.5%						
Subtotal	943	30.7%		643	20.9%		535	17.4%		335	10.9%
Excl. within-category duplicates	911	29.6%		629	20.5%		483	15.7%		326	10.6%

ELEVATION			RESOURCES			MISCELLANEOUS			NO GEOGRAPHY		
Characteristic	*N*	*%*	*Characteristic*	*N*	*%*	*Characteristic*	*N*	*%*	*Characteristic*	*N*	*%*
Mountain	173	60.1%	Salt	24	27.0%	Named after	42	47.2%	Miscellaneous	375	26.7%
Elevation	34	11.8%	Reed	18	20.2%	"ithi"	20	22.5%	-ingen	315	22.5%
Stone / Cliff	34	11.8%	Gravel	14	15.7%	Border region	10	11.2%	-heim	256	18.2%
Hill	26	9.0%	Clay	9	10.1%	Burial	3	3.4%	-dorf	177	12.6%
Promontory	10	3.5%	Grain	4	4.5%	Grind	2	2.2%	-hausen	168	12.0%
Drainage divide	4	1.4%	Hunting	3	3.4%	Rain	2	2.2%	-weiler	52	3.7%
Foothills	4	1.4%	Leek / Onion	3	3.4%	Winter	2	2.2%	-acum	38	2.7%
Ridge	2	0.7%	Honey	2	2.2%	Unique cases	8	9.0%	-beuren	22	1.6%
Mudflow	1	0.3%	Iron	2	2.2%						
			Unique cases	10	11.2%						
Subtotal	288	9.4%		89	2.9%		89	2.9%		1,403	45.6%
Excl. within-category duplicates	288	9.4%		88	2.9%		89	2.9%		1,375	44.7%

Notes: Subtotal percentages are shares of 3,705 toponyms in the dataset.

Table 3. Descriptive statistics

Country	Variable	Mean	Std. dev.	Min.	Max.	N
Full sample	Log population (1910)	7.90	1.30	2.30	14.55	3,705
	Log urban potential (1910)	10.75	1.25	5.08	19.24	3,705
	First record (year)	1098	235	-40	1921	3,705
	City rights awarded (year)	1520	320	50	2019	2,093
Austria	Log population (1910)	8.80	1.04	6.60	14.55	125
	Log urban potential (1910)	11.00	1.01	9.07	16.62	125
	First record (year)	1034	250	50	1352	125
	City rights awarded (year)	1601	339	1014	2001	111
Belgium	Log population (1910)	8.12	0.57	7.71	8.52	2
	Log urban potential (1910)	10.83	0.56	10.43	11.22	2
	First record (year)	890	342	648	1131	2
	City rights awarded (year)	1175	247	1000	1350	2
Czech Republic	Log population (1910)	9.76	0.62	8.56	10.87	39
	Log urban potential (1910)	12.14	0.70	10.97	13.48	39
	First record (year)	1224	130	993	1506	39
	City rights awarded (year)	1471	265	1223	1952	39
France	Log population (1910)	8.43	1.36	6.18	12.09	42
	Log urban potential (1910)	11.24	1.37	8.95	15.04	42
	First record (year)	985	272	350	1629	42
	City rights awarded (year)	1284	136	1150	1580	15
Germany	Log population (1910)	7.74	1.26	2.30	13.74	3,187
	Log urban potential (1910)	10.67	1.27	5.08	19.24	3,187
	First record (year)	1099	233	-40	1921	3,187
	City rights awarded (year)	1531	325	50	2017	1,735
Italy	Log population (1910)	8.77	0.99	7.75	10.32	8
	Log urban potential (1910)	11.02	0.98	9.91	12.53	8
	First record (year)	817	349	150	1256	8
	City rights awarded (year)	1345	125	1200	1536	5
Latvia	Log population (1910)	9.97		9.97	9.97	1
	Log urban potential (1910)	11.40		11.40	11.40	1
	First record (year)	1253		1253	1253	1
	City rights awarded (year)	1328		1328	1328	1
Lithuania	Log population (1910)	8.37	0.95	7.07	10.72	17
	Log urban potential (1910)	11.07	0.94	9.81	13.44	17
	First record (year)	997	210	634	1274	17
	City rights awarded (year)	1358	306	952	1907	10
Poland	Log population (1910)	9.43	0.87	6.80	13.15	131
	Log urban potential (1910)	11.43	0.92	8.54	15.24	131
	First record (year)	1256	119	970	1645	131
	City rights awarded (year)	1376	196	1187	1959	124
Russia	Log population (1910)	9.71	1.35	8.57	12.41	8
	Log urban potential (1910)	11.23	1.35	10.07	13.98	8
	First record (year)	1358	111	1255	1580	8
	City rights awarded (year)	1587	187	1286	1772	8
Switzerland	Log population (1910)	8.37	0.94	6.15	12.28	145
	Log urban potential (1910)	11.08	0.93	8.94	15.14	145
	First record (year)	996	253	150	1788	145
	City rights awarded (year)	1453	327	1045	2019	43

Table 4. Toponyms and Zipf's Law

		Panel A: All geography			Panel B: Excl. Structures		
I. Full sample	*Min. frequency*	*Coefficient (s.e.)*	R^2	*# char.*	*Coefficient (s.e.)*	R^2	*#char.*
	≥0	-0.648 (0.012)	0.949	168	-0.635 (0.014)	0.942	124
	≥5	-0.850 (0.024)	0.948	68	-0.848 (0.028)	0.947	52
	≥10	-0.961 (0.036)	0.942	44	-0.957 (0.037)	0.950	36
	≥15	-1.043 (0.044)	0.944	35	-1.026 (0.046)	0.947	28
	≥20	-1.080 (0.051)	0.939	30	-1.067 (0.058)	0.940	23
	≥25	-1.123 (0.066)	0.926	24	-1.107 (0.072)	0.930	19
	≥30	-1.195 (0.086)	0.914	19	-1.152 (0.096)	0.911	15
II. German sample	*Min. frequency*	*Coefficient (s.e.)*	R^2	*# char.*	*Coefficient (s.e.)*	R^2	*#char.*
	≥0	-0.664 (0.013)	0.947	152	-0.652 (0.015)	0.940	114
	≥5	-0.866 (0.027)	0.944	62	-0.867 (0.030)	0.947	49
	≥10	-1.008 (0.038)	0.948	40	-0.986 (0.040)	0.949	33
	≥15	-1.080 (0.046)	0.948	32	-1.051 (0.053)	0.942	25
	≥20	-1.149 (0.058)	0.942	25	-1.118 (0.067)	0.936	20
	≥25	-1.180 (0.067)	0.936	22	-1.142 (0.075)	0.931	18
	≥30	-1.213 (0.085)	0.922	18	-1.162 (0.103)	0.906	14

Note: The coefficients show the OLS estimates of the log rank with the log frequency of geographic characteristics, i.e., ζ in equation (3). Standard errors in parentheses.

Table 5. Zipf and early geographic advantage

		Panel A: No controls			Panel B: Incl. bicentennial FEs		
I. Full sample	*Maximum gap*	*Coefficient (s.e.)*	R^2	*N*	*Coefficient (s.e.)*	R^2	*N*
	No restrictions	0.830 (0.004)	0.975	1,269	0.896 (0.004)	0.987	1,269
	≤700	0.854 (0.003)	0.983	1,219	0.903 (0.004)	0.989	1,219
	≤600	0.860 (0.003)	0.984	1,189	0.906 (0.004)	0.990	1,189
	≤500	0.876 (0.003)	0.986	1,085	0.913 (0.004)	0.990	1,085
	≤400	0.905 (0.003)	0.988	925	0.925 (0.004)	0.990	925
	≤300	0.933 (0.003)	0.991	790	0.944 (0.004)	0.991	790
	≤200	0.965 (0.004)	0.991	601	0.970 (0.004)	0.992	601
	≤100	1.000 (0.006)	0.987	340	1.002 (0.006)	0.988	340
II. German sample							
	No restrictions	0.881 (0.004)	0.979	1,034	0.934 (0.005)	0.988	1,034
	≤700	0.900 (0.003)	0.986	1,000	0.943 (0.004)	0.990	1,000
	≤600	0.906 (0.003)	0.986	975	0.946 (0.004)	0.990	975
	≤500	0.920 (0.004)	0.987	883	0.953 (0.004)	0.990	883
	≤400	0.946 (0.004)	0.988	735	0.965 (0.004)	0.989	735
	≤300	0.977 (0.004)	0.990	610	0.985 (0.005)	0.990	610
	≤200	1.008 (0.005)	0.989	455	1.011 (0.005)	0.989	455
	≤100	1.052 (0.008)	0.984	249	1.052 (0.009)	0.984	249

Note: The coefficients show the OLS estimates of the log rank with the log year gap of city rights received and first record mention of the toponym. Panel A shows the estimates of η for equation (6) without additional controls; panel B includes bicentennial dummies starting from year 1. Intercept estimates not presented for brevity. Standard errors in parentheses.

Table 6. Toponyms and urban development: Main results

	(1) Water		(2) Structures		(3) Vegetation		(4) Land		(5) Elevation		(6) Resources		(7) All	
WATER:														
- Stream	-0.007	(0.01)											-0.001	(0.01)
- River	**0.034**	**(0.01)*****											**0.036**	**(0.01)*****
- Floodplain	0.010	(0.01)											0.012	(0.01)
- Ford	**0.064**	**(0.02)*****											**0.066**	**(0.02)*****
- Source	-0.033	(0.02)											-0.034	(0.02)
- Lake	-0.009	(0.03)											-0.014	(0.03)
- Water	0.031	(0.03)											0.035	(0.04)
- River mouth	0.040	(0.03)											0.046	(0.03)
- River island	0.021	(0.02)											0.022	(0.02)
STRUCTURES:														
- Castle			0.008	(0.01)									0.006	(0.01)
- Church			**0.056**	**(0.01)*****									**0.061**	**(0.02)*****
- Farm estate			-0.008	(0.02)									-0.009	(0.02)
- Monastery			0.023	(0.02)									0.020	(0.02)
- Bridge			0.029	(0.03)									0.028	(0.03)
- Fence			**0.059**	**(0.03)***									**0.059**	**(0.03)***
- Mill			0.011	(0.03)									0.013	(0.04)
VEGETATION:														
- Tree					0.018	(0.01)							0.022	(0.01)
- Forest					-0.036	(0.05)							-0.029	(0.05)
- Clearing					**0.045**	**(0.02)*****							**0.054**	**(0.02)*****
- Hedge					**-0.073**	**(0.02)*****							**-0.069**	**(0.02)*****
- Bush					0.010	(0.03)							0.010	(0.03)
- Heath					-0.441	(0.43)							-0.431	(0.44)
LAND:														
- Field							-0.012	(0.01)					-0.009	(0.02)
- Swamp							-0.022	(0.02)					-0.018	(0.02)
- Bog							0.011	(0.03)					0.007	(0.03)
- Land							0.024	(0.03)					0.021	(0.03)
- Valley							0.016	(0.04)					0.015	(0.04)
ELEVATION:														
- Mountain									0.013	(0.01)			0.014	(0.01)
- Elevation									0.006	(0.02)			0.006	(0.02)
- Stone/cliff									0.007	(0.02)			0.005	(0.02)
- Hill									0.037	(0.02)			**0.041**	**(0.02)***

(continued on next page)

(continued from previous page)

RESOURCES:														
- Salt											0.013	(0.03)	0.014	(0.03)
- Reed											0.027	(0.02)	0.024	(0.02)
CONTROLS:														
No geography	-0.016	(0.01)***	-0.015	(0.01)***	-0.021	(0.01)***	-0.018	(0.00)***	-0.017	(0.00)***	-0.018	(0.00)***	-0.015	(0.01)*
log(UP)	0.949	(0.03)***	0.950	(0.03)***	0.951	(0.03)***	0.950	(0.03)***	0.951	(0.03)***	0.950	(0.03)***	0.948	(0.03)***
First record	-0.077	(0.04)*	-0.081	(0.04)*	-0.072	(0.04)*	-0.081	(0.04)*	-0.079	(0.04)**	-0.078	(0.04)*	-0.073	(0.04)*
City rights	0.130	(0.03)***	0.132	(0.03)***	0.128	(0.03)***	0.131	(0.03)***	0.131	(0.03)***	0.131	(0.03)***	0.130	(0.03)***
Country FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lat & Lon	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Observations	3,705		3,705		3,705		3,705		3,705		3,705		3,705	
Adjusted R^2	0.969		0.969		0.970		0.969		0.970		0.969		0.969	

Notes: Dependent variable is log(1910 population). Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Parameter estimates for the Miscellaneous category, characteristics with <15 observations, and the constant are not displayed for brevity.

Table 7. Toponyms and urban development: Sensitivity analyses

	(1) Excl. multiple interpretations		(2) Excl. no interpretation		(3) Excl. mixed meanings		(4) Excl. imprecise first record		(5) Column (1)-(4) combined	
WATER:										
- Stream	0.005	(0.01)	0.005	(0.02)	-0.004	(0.02)	-0.003	(0.01)	-0.005	(0.02)
- River	**0.038**	**(0.01)*** **	**0.040**	**(0.01)*** **	**0.038**	**(0.02)** **	**0.032**	**(0.01)*** **	**0.032**	**(0.02)***
- Floodplain	0.020	(0.02)	0.016	(0.02)	0.015	(0.02)	0.010	(0.01)	0.015	(0.02)
- Ford	**0.074**	**(0.03)*** **	**0.073**	**(0.03)*** **	**0.071**	**(0.03)** **	**0.057**	**(0.02)*** **	**0.064**	**(0.03)** **
- Source	-0.028	(0.02)	-0.030	(0.02)	-0.036	(0.02)	-0.035	(0.02)	-0.034	(0.02)
- Lake	-0.007	(0.03)	-0.015	(0.03)	-0.013	(0.03)	-0.007	(0.03)	0.000	(0.03)
- Water	0.042	(0.04)	0.042	(0.04)	0.057	(0.05)	-0.013	(0.03)	0.008	(0.04)
- River mouth	**0.056**	**(0.03)***	0.053	(0.03)	0.054	(0.04)	0.039	(0.03)	0.052	(0.03)
- River island	0.027	(0.02)	0.024	(0.02)	0.015	(0.03)	0.029	(0.02)	0.025	(0.02)
STRUCTURES:										
- Castle	0.010	(0.01)	0.009	(0.01)	0.002	(0.01)	0.007	(0.01)	0.006	(0.01)
- Church	**0.065**	**(0.02)*** **	**0.066**	**(0.02)*** **	**0.068**	**(0.02)*** **	**0.057**	**(0.02)*** **	**0.064**	**(0.02)*** **
- Farm estate	-0.011	(0.02)	-0.004	(0.02)	0.007	(0.02)	-0.008	(0.02)	0.002	(0.02)
- Monastery	0.024	(0.02)	0.024	(0.02)	0.013	(0.02)	0.024	(0.02)	0.018	(0.02)
- Bridge	0.032	(0.03)	0.033	(0.03)	0.026	(0.03)	0.029	(0.03)	0.029	(0.03)
- Fence	**0.067**	**(0.03)** **	**0.066**	**(0.03)***	**0.062**	**(0.04)***	**0.052**	**(0.03)***	0.060	(0.04)
- Mill	0.015	(0.04)	0.015	(0.04)	-0.002	(0.05)	-0.002	(0.03)	0.003	(0.05)
VEGETATION:										
- Tree	**0.028**	**(0.01)***	0.024	(0.02)	0.011	(0.02)	**0.024**	**(0.01)***	0.017	(0.02)
- Forest	-0.028	(0.05)	-0.024	(0.04)	-0.033	(0.05)	-0.042	(0.05)	-0.048	(0.06)
- Clearing	**0.060**	**(0.02)*** **	**0.060**	**(0.02)*** **	**0.058**	**(0.02)** **	**0.054**	**(0.02)*** **	**0.057**	**(0.02)** **
- Hedge	**-0.063**	**(0.02)*** **	**-0.065**	**(0.02)*** **	**-0.086**	**(0.02)*** **	**-0.067**	**(0.02)*** **	**-0.082**	**(0.02)*** **
- Bush	0.007	(0.03)	0.009	(0.03)	0.003	(0.06)	0.009	(0.03)	-0.008	(0.06)
- Heath	-0.423	(0.43)	-0.425	(0.43)	-0.456	(0.46)	-0.449	(0.45)	-0.477	(0.47)
LAND:										
- Field	-0.006	(0.02)	-0.005	(0.02)	-0.007	(0.02)	-0.008	(0.02)	-0.006	(0.02)
- Swamp	-0.014	(0.02)	-0.013	(0.02)	-0.022	(0.03)	-0.014	(0.02)	-0.019	(0.03)
- Bog	0.006	(0.03)	0.013	(0.03)	0.005	(0.04)	-0.011	(0.03)	-0.027	(0.04)
- Land	0.027	(0.03)	0.026	(0.03)	0.015	(0.03)	0.012	(0.03)	0.013	(0.03)
- Valley	0.020	(0.04)	0.019	(0.04)	0.012	(0.04)	0.020	(0.04)	0.025	(0.05)
ELEVATION:										
- Mountain	0.020	(0.01)	0.017	(0.01)	0.012	(0.02)	0.013	(0.01)	0.016	(0.02)
- Elevation	0.012	(0.02)	0.011	(0.02)	0.042	(0.03)	0.007	(0.02)	0.048	(0.03)
- Stone/cliff	0.010	(0.02)	0.009	(0.03)	0.015	(0.03)	0.008	(0.02)	0.021	(0.03)
- Hill	**0.045**	**(0.02)***	**0.044**	**(0.02)***	**0.055**	**(0.03)** **	**0.045**	**(0.02)***	**0.063**	**(0.03)** **

(continued on next page)

(continued from previous page)

RESOURCES:										
- Salt	0.017	(0.03)	0.018	(0.03)	0.021	(0.03)	0.009	(0.03)	0.017	(0.03)
- Reed	0.023	(0.02)	0.027	(0.02)	0.004	(0.03)	0.032	(0.02)	0.004	(0.02)
CONTROLS:										
No geography	-0.007	(0.01)	-0.011	(0.01)	-0.020	(0.01)	**-0.013**	**(0.01)***	-0.014	(0.01)
log(UP)	**0.946**	**(0.03)*****	**0.944**	**(0.03)*****	**0.939**	**(0.03)*****	**0.962**	**(0.02)*****	**0.954**	**(0.03)*****
First record	**-0.073**	**(0.04)***	**-0.076**	**(0.04)***	**-0.087**	**(0.04)****	**-0.050**	**(0.03)***	-0.056	(0.03)
City rights	**0.135**	**(0.04)*****	**0.134**	**(0.04)*****	**0.139**	**(0.04)*****	**0.109**	**(0.03)*****	**0.119**	**(0.04)*****
Country FE	Yes		Yes		Yes		Yes		Yes	
Lat & Lon	Yes		Yes		Yes		Yes		Yes	
Observations	3,523		3,485		3,198		3,556		2,965	
Adjusted R^2	0.968		0.968		0.966		0.976		0.973	

Notes: Dependent variable is log(1910 population). Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Column (1) excludes toponyms with multiple possible interpretations. Column (2) excludes toponyms if there is no consensus, preferred or any interpretation. Column (3) excludes toponyms containing both geographic and non-geographic meanings. Column (4) excludes places of which the first-record year is imprecise. Column (5) combines the restrictions from columns (1) through (4). Parameter estimates for the Miscellaneous category, characteristics with <15 observations, and the constant are not displayed for brevity.

Table 8. Geography and early geographical advantage: Baseline and sensitivity results

	(1) Full sample		(2) Excl. multiple interpretations		(3) Excl. no interpretation		(4) Excl. mixed characteristics		(5) Excl. imprecise first-record year		(6) Column (2)-(5) combined	
WATER:												
- Stream	**0.217**	**(0.09)****	**0.233**	**(0.09)*****	**0.261**	**(0.09)*****	**0.217**	**(0.09)****	**0.276**	**(0.10)*****	**0.276**	**(0.11)****
- River	-0.063	(0.08)	-0.060	(0.09)	-0.027	(0.08)	-0.071	(0.08)	-0.019	(0.10)	-0.030	(0.10)
- Floodplain	-0.178	(0.17)	-0.184	(0.18)	-0.154	(0.17)	-0.156	(0.18)	-0.126	(0.19)	-0.110	(0.20)
- Ford	-0.080	(0.21)	0.056	(0.16)	-0.049	(0.21)	-0.118	(0.21)	-0.090	(0.22)	-0.007	(0.18)
- Source	**0.360**	**(0.22)***	**0.447**	**(0.21)****	**0.404**	**(0.22)***	**0.395**	**(0.23)***	**0.369**	**(0.22)***	**0.487**	**(0.23)****
- Lake	-0.326	(0.21)	-0.393	(0.21)*	-0.306	(0.21)	**-0.377**	**(0.22)***	-0.201	(0.21)	**-0.367**	**(0.18)****
- Water	-0.126	(0.26)	-0.271	(0.29)	-0.097	(0.27)	-0.186	(0.31)	-0.105	(0.30)	-0.292	(0.39)
- River mouth	-0.163	(0.30)	-0.170	(0.30)	-0.117	(0.30)	-0.152	(0.31)	-0.123	(0.31)	-0.111	(0.32)
- River island	**0.399**	**(0.21)***	**0.395**	**(0.21)***	**0.452**	**(0.22)****	**0.395**	**(0.21)***	**0.497**	**(0.24)****	**0.467**	**(0.24)***
STRUCTURES:												
- Castle	**-0.330**	**(0.09)*****	**-0.325**	**(0.09)*****	**-0.286**	**(0.09)*****	**-0.353**	**(0.09)*****	**-0.285**	**(0.10)*****	**-0.320**	**(0.10)*****
- Church	**0.257**	**(0.14)***	**0.250**	**(0.14)***	**0.306**	**(0.14)****	**0.264**	**(0.14)***	**0.358**	**(0.15)****	**0.353**	**(0.16)****
- Farm estate	**0.407**	**(0.13)*****	**0.461**	**(0.13)*****	**0.456**	**(0.14)*****	**0.409**	**(0.14)*****	**0.533**	**(0.16)*****	**0.596**	**(0.15)*****
- Monastery	0.031	(0.22)	0.020	(0.22)	0.076	(0.22)	0.027	(0.22)	0.074	(0.22)	0.050	(0.23)
- Bridge	**-0.440**	**(0.25)***	-0.335	(0.24)	-0.407	(0.25)	-0.433	(0.26)	**-0.516**	**(0.28)***	-0.396	(0.27)
- Fence	0.152	(0.20)	0.160	(0.21)	0.188	(0.21)	0.138	(0.20)	0.177	(0.21)	0.154	(0.21)
- Mill	-0.232	(0.26)	-0.230	(0.26)	-0.218	(0.25)	-0.272	(0.27)	-0.753	(0.49)	-0.751	(0.49)
VEGETATION:												
- Tree	0.086	(0.11)	0.092	(0.12)	0.107	(0.12)	0.093	(0.12)	0.171	(0.13)	0.189	(0.13)
- Forest	0.188	(0.12)	0.183	(0.12)	**0.215**	**(0.12)***	0.170	(0.13)	0.172	(0.14)	0.128	(0.15)
- Clearing	0.209	(0.19)	0.215	(0.19)	0.256	(0.19)	0.207	(0.19)	0.290	(0.21)	0.276	(0.22)
- Hedge	-0.149	(0.30)	-0.155	(0.30)	-0.108	(0.30)	-0.302	(0.33)	-0.224	(0.31)	-0.444	(0.35)
- Bush	0.004	(0.47)	-0.001	(0.48)	0.003	(0.47)	0.070	(0.49)	-0.329	(0.43)	-0.261	(0.54)
- Heath	-0.114	(0.40)	-0.149	(0.42)	-0.112	(0.41)	-0.103	(0.40)	0.014	(0.47)	-0.038	(0.49)
LAND:												
- Field	0.067	(0.12)	0.055	(0.13)	0.106	(0.13)	0.052	(0.13)	0.047	(0.13)	-0.002	(0.14)
- Swamp	**0.299**	**(0.17)***	0.273	(0.17)	**0.332**	**(0.17)***	**0.340**	**(0.18)***	0.266	(0.18)	0.201	(0.19)
- Bog	0.358	(0.44)	0.429	(0.42)	0.446	(0.44)	0.523	(0.50)	0.516	(0.55)	**0.909**	**(0.55)***
- Land	-0.060	(0.28)	-0.076	(0.28)	-0.042	(0.28)	-0.011	(0.29)	-0.013	(0.30)	-0.053	(0.31)
- Valley	0.259	(0.38)	0.262	(0.38)	0.310	(0.38)	0.276	(0.44)	0.208	(0.45)	0.213	(0.54)
ELEVATION:												
- Mountain	-0.187	(0.14)	-0.131	(0.14)	-0.146	(0.14)	-0.189	(0.14)	-0.174	(0.15)	-0.135	(0.15)
- Elevation	0.076	(0.18)	0.084	(0.18)	0.098	(0.18)	0.071	(0.18)	0.038	(0.24)	0.013	(0.24)
- Stone/cliff	-0.269	(0.28)	-0.278	(0.29)	-0.229	(0.29)	-0.285	(0.31)	-0.252	(0.32)	-0.301	(0.35)
- Hill	0.433	(0.41)	0.421	(0.41)	0.453	(0.40)	0.435	(0.41)	0.425	(0.50)	0.398	(0.51)

(continued on next page)

(continued from previous page)

RESOURCES:												
- Salt	0.061	(0.21)	0.051	(0.21)	0.103	(0.22)	0.035	(0.22)	0.193	(0.24)	0.149	(0.26)
- Reed	-0.202	(0.29)	-0.303	(0.30)	-0.177	(0.29)	-0.030	(0.25)	-0.151	(0.32)	-0.046	(0.29)
CONTROLS:												
No geography	**0.163**	**(0.07)****	**0.161**	**(0.07)****	**0.210**	**(0.07)*****	**0.153**	**(0.07)****	**0.202**	**(0.10)***	0.169	(0.11)
Country FE	Yes		Yes		Yes		Yes		Yes		Yes	
Lat & Lon	Yes		Yes		Yes		Yes		Yes		Yes	
Bicentennial FE	Yes		Yes		Yes		Yes		Yes		Yes	
Observations	2,004		1,900		1,884		1,931		1,721		1,597	
Adjusted R^2	0.346		0.344		0.350		0.339		0.351		0.348	

Notes: Dependent variable is the log of the difference between the year that a place received city rights and its first record. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Column (1) reports the baseline for the full sample. Column (2) excludes toponyms with multiple possible interpretations. Column (3) excludes toponyms without a (preferred or consensus-based) interpretation. Column (4) excludes toponyms containing both geographic and non-geographic meanings. Column (5) excludes places of which the first-record year is imprecise. Column (6) combines the restrictions from columns (2) through (5). Parameter estimates for the Miscellaneous category, characteristics with <15 observations, controls and the constant are not displayed for brevity.

Table 9. Interaction between characteristics and early cities

	(1) Main effect		(2) Interaction with pre-1800 city rights		(3) Main effect		(4) Interaction with pre-1800 city rights	
WATER:								
- Stream	**-0.046 (0.01)**	***	**0.115 (0.02)**	***	**-0.052 (0.01)**	***	**0.119 (0.03)**	***
- River	0.017 (0.01)		**0.048 (0.02)**	**	0.011 (0.02)		**0.053 (0.02)**	**
- Floodplain	-0.024 (0.02)		**0.073 (0.03)**	***	-0.019 (0.02)		**0.068 (0.03)**	**
- Ford	**0.066 (0.02)**	***	0.039 (0.04)		**0.065 (0.03)**	**	0.037 (0.04)	
- Source	**-0.078 (0.03)**	***	**0.113 (0.06)**	**	**-0.076 (0.03)**	**	**0.110 (0.06)**	*
- Lake	-0.043 (0.04)		0.058 (0.05)		-0.010 (0.04)		0.009 (0.07)	
- Water	0.019 (0.03)		0.016 (0.07)		-0.017 (0.03)		0.049 (0.09)	
- River mouth	**0.119 (0.03)**	***	-0.064 (0.05)		**0.131 (0.03)**	***	**-0.078 (0.05)**	*
- River island	0.013 (0.06)		0.049 (0.06)		0.012 (0.05)		0.053 (0.06)	
STRUCTURES:								
- Castle	0.022 (0.02)		0.019 (0.02)		0.012 (0.02)		0.027 (0.02)	
- Church	**0.046 (0.02)**	***	0.012 (0.03)		**0.044 (0.02)**	*	0.028 (0.03)	
- Farm estate	**-0.036 (0.02)**	*	0.051 (0.04)		**-0.042 (0.02)**	*	**0.107 (0.03)**	***
- Monastery	0.012 (0.02)		0.027 (0.04)		0.009 (0.02)		0.026 (0.04)	
- Bridge	0.004 (0.05)		0.068 (0.06)		-0.005 (0.05)		0.079 (0.07)	
- Fence	0.061 (0.04)		-0.003 (0.05)		0.066 (0.04)		-0.011 (0.05)	
- Mill	0.036 (0.06)		-0.023 (0.08)		0.006 (0.08)		0.015 (0.09)	
VEGETATION:								
- Tree	0.017 (0.02)		0.023 (0.03)		0.010 (0.02)		0.031 (0.03)	
- Forest	-0.073 (0.07)		0.126 (0.08)		-0.110 (0.10)		0.156 (0.12)	
- Clearing	0.021 (0.02)		**0.083 (0.03)**	***	0.026 (0.03)		**0.078 (0.03)**	***
- Hedge	**-0.107 (0.03)**	***	**0.103 (0.04)**	**	**-0.133 (0.03)**	***	**0.119 (0.05)**	**
- Bush	-0.026 (0.05)		0.073 (0.06)		-0.018 (0.08)		0.020 (0.09)	
- Heath	-0.553 (0.56)		0.481 (0.55)		-0.559 (0.56)		0.416 (0.55)	
ELEVATION:								
- Mountain	-0.016 (0.02)		**0.069 (0.03)**	***	-0.016 (0.02)		**0.065 (0.03)**	**
- Elevation	-0.006 (0.03)		**0.072 (0.04)**	*	0.045 (0.04)		0.041 (0.04)	
- Stone / cliff	-0.040 (0.04)		**0.114 (0.05)**	**	-0.035 (0.04)		**0.130 (0.05)**	**
- Hill	0.008 (0.03)		0.060 (0.05)		0.032 (0.03)		0.041 (0.06)	
LAND:								
- Field	**-0.043 (0.02)**	**	**0.098 (0.03)**	***	**-0.047 (0.02)**	**	**0.103 (0.04)**	***
- Swamp	-0.042 (0.03)		0.040 (0.04)		**-0.053 (0.03)**	*	0.067 (0.05)	
- Bog	-0.025 (0.03)		**0.144 (0.07)**	**	**-0.058 (0.03)**	*	**0.000 (0.00)**	***
- Land	-0.006 (0.04)		0.039 (0.06)		-0.007 (0.04)		0.010 (0.05)	
- Valley	0.030 (0.04)		-0.055 (0.06)		0.021 (0.05)		-0.041 (0.07)	
RESOURCES:								
- Salt	0.008 (0.06)		0.034 (0.06)		-0.012 (0.07)		0.075 (0.08)	
- Reed	0.039 (0.03)		-0.006 (0.05)		0.015 (0.04)		0.018 (0.06)	
CONTROLS:								
No geography	**-0.054 (0.01)**	***	**0.091 (0.02)**	***	**-0.057 (0.01)**	***	**0.097 (0.02)**	***
log(UP)	**0.958 (0.02)**	***			**0.961 (0.03)**	***		
First record	**-0.082 (0.04)**	***			-0.062 (0.04)			
Country FE		Yes				Yes		
Lat & Lon		Yes				Yes		
Observations		3,705				2,965		
Adjusted R^2		0.969				0.972		

Notes: Dependent variable is log(1910 population). Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Column (1) and (2) include the full sample. Column (3) and (4) exclude toponyms with multiple possible interpretations, no (preferred or consensus-based) interpretation, toponyms containing both geographic and non-geographic meanings, and places of which the first-record year is imprecise. Parameter estimates for the Miscellaneous category, characteristics with <15 observations, other controls and the constant are not displayed for brevity.

Appendix A

A.1. Toponym coding

Each entry in Niemeyer (2012) is categorized into five parts, (I) a summary of the place including its current population, legal status as municipality or city, and a description of the location of the place that is especially useful to identify which place is actually meant in case of a common toponym. The year in which city rights were awarded is not always provided. In this case the legal status of the place provides guidance to whether the place had obtained city rights in the past. In case of missing years, we add them manually and typically those were taken from the place's official municipal website. (II) provides an overview of the names of the place over time, as well as the year of the recorded first mention. (III) provides the toponomy, that is, the origin and meaning of the toponym of interest. On occasion, the precise meaning is not immediately clear, often due to similar sounding suffixes i.e. -berg (mountain) and -burg (castle), or because it is a rather unique name (such as Dortmund) or variant of a toponym (Dörpen, instead of the common *-dorf*). As such this category provides more of an interpretation of the most plausible meaning by today's research on the toponym, rather than a definite answer. (IV) provides similar toponyms in this compendium, which is useful for cross-verification of toponym meanings. Lastly, (V) shows the source of this entry, as the book is a collection summaries of research articles in toponomy by various authors

To translate an example page, the entry Dornstadt in Niemeyer (2012, page 135) reads 'I. municipality and *Vereinbarte Verwaltungsgemeinschaft* in the Alb-Donau-district, 13,185 population, 8 km north-north-west of Ulm and ca. 42km south-south-east of Göppingen [...] II. 1225 Dorneconstat, 1330 Dorgenstat, 1334 Dorgunstat, 1643 Dornstat III. The name is to be interpreted as 'habitation on the thorny place' and belongs to middle high-German *dornec* 'thorny' and the basic, old high-German, middle high-German *stat* 'site, place, habitation'. The basis on lower high-German is secondary, the dropping out of the *g* is Alemannic. Eponymous were the old thorny hedges of the Muschelkalk area of the Heckengäu, which grow on the clearance cairn of the field border.

We assign this place to the 'hedges' type on the 'suffix-level' (assign a 1 to the *vegetation_hedges* dummy), with the additional description 'thorny' as a reference to a hedge in this case, so it does not fit into the traditional suffix/prefix dichotomy. We also assign a 1 to the *no_geography_stadt* dummy, so this toponym is an example of a place with both a geographic and non-geographic meaning in its original name.

Similarly, the next entry Dornstetten has the same meaning, and is written and coded in the dataset as such. The next entry Dörpen is rather short, which usually indicates that the interpretation is straightforward. Here, 'II. 890 Dorpun [...] III. the name consists of the simplex, old-Saxon *thorp*, middle lower-German *dorp*, in the dative plural case, *-dorf*'. So, the *no_geography_dorf* dummy is coded 1. The next two entries, however, are not as straightforward. The city Dorsten reads 'II. around 900 Durstinon [...] III. Lower German *dorst*, today narrowed down to 'thirst', like Latin *torrere* 'drying, parching', and show high-German parching and dry, a general meaning in the semantic area of 'drying'. The suffix *-ina* is, in

settlement names, often interpreted as a 'waterbody', can also be understood as a diminutive suffix or can be used to construct adjectives. Motive for the naming could be, for the preservation of certain fruits, indispensable *darre* (drying facility). The word with -st- also leaves the possibility of an 'old European' water body name. The unusual retention of the -o- instead of the -i- in the following syllable to be expected umlaut -ö- has parallels to the neighboring Netherlands'. For this case, we do not assign Dorsten to any type, as it is unclear and no favorite or likely interpretation is provided in the entry. Furthermore, it is mentioned that Dorsten became a fortified city in 1251, which is unclear whether the city got simply fortified in 1251, or received the city rights in 1251 with the typical rights to build walls. We double-checked these particular cases with official (online) sources where it was unclear, and indeed confirm the receipt of city rights in 1251. In the end, we do not assign this place either a geographic or non-geographic meaning. In this case, we remark this entry as having 'nothing assigned', which will be used in the robustness exercises. However, we keep this entry as it is still useful in the calculation of a Zipf's law for natural advantages in section 4.2. As the first record is said to be around 900 as 'Durstinon', we also set the year of the first record at 900, ignoring the 'around' approximation in this case.

Lastly, Dortmund is 'II. around 890 Throtmanni [...] III. After many, partly amateurish attempt to interpret the name, P. Derks 1987 confirmed a philologically justifiable explanation, which was suggested already in 1919: Old Estonian *throtu* and old high-German *drozza* 'throat, gullet, maw' (cf lower high-German choke) [...] The basic word *-menni* is in many settlement names documented as waters. The -d at the end came about, as documented, only later as a reference to *mund* 'protection'. Therefore, the place name Dortmund is a description of a throat like shape of terrain (cf. hollow throat) flowing stream'. We therefore assign Dortmund to the 'stream' type suffix-level (the variable *water_stream* is coded 1). Again, it is reference to a stream in this case, so it does not fit into the traditional suffix/prefix dichotomy very often found with stream places that end in *-bach* or *-ach* most of the time.

A.2. Toponym aggregation

Our coding of toponyms reveals a large number of characteristics that require a systematic ordering for empirical analysis. In the dataset provided in the online repository, we identify 168 characteristics and sort them into 6 geographical categories (Water, Structures, Vegetation, Elevation, Land, Resources) for the ease of exposition, and 1 no-geography category. Table A1 presents the complete list of these characteristics sorted by categories, along with their characteristics, the number of observations containing these characteristics, their common toponym(s) and additional notes.

<< TABLE A1 ABOUT HERE >>

A.3. Changing names over time

It is often the case that the toponym of the first record is not exactly the same as the toponym of today, due to centuries of changes driven simply by changes in language, or mistranscriptions accumulating over time. For instance, Dortmund went from Throtmanni (890) to Throtmennia (947) to Drutmunne (1074). Although the interpretation of the meaning did not change in this specific example, some unresolvable ambiguity could still be introduced through mistranscriptions. For example, Friedberg (Bavaria) was first mentioned as ciutatem Fridberch (1264). However, a castle had already been built there in 1260, with peace in mind, to secure the Bavarian border to Augsburg. It is therefore more likely that the name of the place actually means 'peace castle', and not 'peace mountain'.

This brings us to castles being often found on mountains, such that *-berg* (mountain) and -burg (castle) were often interchangeable, and hence the true meaning of a place may be ambiguous and may often change back and forth over the centuries. In the case of Friedberg (Bavaria), it is clear which was meant. As for other *-berg*/*-burg* places, the book often describes which of the two is actually meant, so that it is not often the case that we have to assign neither of the two due to ambiguity. To further reduce ambiguity, we did additional desk research to determine whether castles and/or mountains had been historically present in the area.[26] Similar to mistranscription is that the people themselves over time misinterpret the original meaning of the place. For example, Neuenhagen bei Berlin was originally *nyenhoue* (1367) 'new farm', which somehow changed over time to 'new hedge'. We only consider the very first recorded mention and therefore always assign the very first mentioned characteristic.

Note that we do not count the change of Gladbach to Mönchengladbach and that of another Gladbach to Bergisch Gladbach as a change in their name, as these clarification prefixes were added to distinguish the two places. This is the case when we have places with the same name in close proximity to each other. The other typical case is when two places used to be one, but for some reason drifted apart, split and/or became significant places on their own to warrant the distinguishing prefix. For instance, the typical distinguishing prefixes are about their respective direction to each other (west, east, etc.), their respective age difference (old, new, etc.), their respective altitude (high, below, etc.), or simply a geographic characteristic (church, castle, mountain). Not adding more characteristics over time would make it consistent with other places where additional distinguishing prefixes were not added.

A unique case is Hirschaid, first mentioned as *Hirzheide* (1097) and translated as 'deer heath', it is likely not the first toponym, but rather the first recorded one, as heaths are not the natural habitat of deers. That is, we never really know if the first recorded mention is the first name given to the place. However, the authors of each entry typically discuss the plausibility of the meaning, e.g., if there is actually a mountain around if the place is named after a mountain, so that the name makes sense. Even if the first recorded mention is different from the first name given, we almost always have a plausible (meaning of the) name.

[26] For a complete list of castles, see alleburgen.de.

Table A1. Toponym categories and characteristics

Characteristic	Freq.	Describing base word	Notes
CATEGORY: WATER			
stream	387	A-, -ach/aha, -bach/beck/beke	Small, flowing, natural water body, often not navigable.
river	248	not applicable	Large, flowing, natural water body, likely more navigable than streams. Toponyms often refer to the river and the hydronyms of rivers are typically much older and retain even if the name is from a different language family. There are no typical, obvious and Germanic river suffixed. The hydronyms themselves often translate to 'move, flow' or other water body references.
floodplain	90	-au/aue/ouw, -grim, -lage, -oog, -ege	Refers to the land adjacent to a water body that is regularly flooded, leaving nutrients behind for agriculture. The most common suffix -au/ouw is very similar to the Slavic suffix -ov/ow (belongs to, place of) which makes careful etymological evaluation necessary here.
ford	38	brod, -förde/furt, -wat, -wedel	Shallow part of a river considered safe to cross, thus making a natural bridge.
source	37	-born/bronn/brunn, -font	Spring or source typically provided by ground water.
lake	29	-see	Non-flowing, natural water body.
water	20	-apa, -ahwa, isa, -opa, -phe	General references to any water in the toponym are grouped here.
river_mouth	17	-mund/gmund	Crossing of two rivers or where a river flows into a different water body.
river_island	15	-werder/werth	Ait (river island) or land between two rivers, e.g. Werder Bremen.
bath	7	Aquae, Bad	References to baths. Not to be confused with the common modern disambiguation Bad often added in the 19-20th centuries to indicate a bath, hot or cold spring in modern times – which we therefore ignored.
marsh	7	marsch	References to marshes (flooded open areas).
port	6	-hafen/haven	No additional notes.
weir	6	-wehr	No additional notes.
boat	5	-fähre, -floss, -kahn, -plau	Various references to boats.
canal	4	-fleet	Mostly references to canals. -fleet are navigable canals regulated by tides.
dam	3	-damm	No additional notes.
fjoerd	3		Narrow inlet into the land formed by glacier movements.
hithe	3	-hude	No direct translation; equivalent to hithe places, i.e., landing places for smaller boats on a water body, and for wood storage.
island	3		No additional notes.
sluice	2	Arke, Schleuse, -siel	No additional notes.
pool	2	-pfühl, -tümpel	As compared to ponds, pools dry out i.e. are temporary.
bay	1	ham	References to bays. Only entry is Hamburg.
big_island	1		Big island in its name, not comparable to a simple island. Only entry is Fehmarn.
dune	1	-lund	The only entry is Schafflund.
groyne	1	Buhne	Dam in a water body to regulate the flow or to prevent erosion.
lagoon	1	Haff/mar	Inner coast waters, or lagoon. Only entry is Mohrungen (Morag).
logs_on_water	1	Specke	The only entry is Haselünne.
pond	1	-teich	As compared to pools, ponds do not typically dry out. Only entry is Rybnyk
portage	1	-walk	The only entry is Pasewalk.
river_arm	1	twist	A diverting section of a river/anabranch. The only entry is Weilerswist, first mentioned as Swist in 1180.
roadstead	1	Reede	The only entry is Rhede.
strait	1	-belt, -sund	The only entry is Stralsund.
CATEGORY: STRUCTURES			
castle	276	-berg/burg, -briga/brica, -dunum, -eck/egg, -grad/gros, hain, -stein	Etymological evaluation is required to distinguish between castle (Burg) and mountain (Berg); -eck and -stein is another case in point due to the proclivity of castles being built on elevated positions.
church	76	-kirch/kirchen	All references to churches, distinct from chapels.
farm_estate	67	-hof/hofen	Farm, smallholding estate, common in Bavaria. Many different meanings are possible such that a clear categorization towards the no-geographic meaning toponyms is not straightforward.
monastery	45	-münster, -zell/zelle	Can also fulfill various services such as education (monastic schools) or copying texts.
bridge	25	-bruck/brück/brugg, -specke	-specke refers to a log bridge. A bridge name often accompanies the name of the water body (e.g., Saarbrücken; Saar river) or a person (e.g., Erndtebrück; bridge of Irmingard) or describes the type of bridge (e.g., Delbrück; plank bridge).

fence	24	-brüel, -esch, -glinde, -ham, -hegel, -hürde, -reke/rike, -tun/zaun, -wik/wiek	Typically described as Einfreidung, a (possibly) man-made fence/enclosure (compare and contrast with a hedge, which may be a natural fence).
mill	20	Mühl-/mul-, Quer-	Often prefixes of different toponyms.
roman_castra	11		References to Roman forts.
outlook_tower	9	-turm, -wecht/wacht	All references to outlooks/vantage points.
walls	9	-macher, -mauer	All references to bulwarks, palisades, and walls.
chapel	7	Bed-, cappeln, Kappel	All references to chapels (small churches)/a small house of prayer
court	7	-dingel, gericht	No additional notes.
barn	6	-scheune, -schuppen	No additional notes.
market	6	-markt	No additional notes.
pole	6	-pfahl, pflock	Could also refer to wooden walls.
road_path	6	-pfad, -stig, -strasse	All references to boardwalks, pathways, roads, streets, etc.
hospital	4	-spiez	No additional notes.
warehouse	3		No additional notes.
drying_place	2	darre	No additional notes.
forestry	2		All references to forestry, which is considered a man-made activity distinct from toponyms referring to forests/trees.
iron_works	2	-hütte	All references to iron works.
shooting_place	2		No additional notes.
shop	2	-kauf	No additional notes.
smithery	2	-schmiede/schmitt	No additional notes.
tavern	2	taverna	All references to inns or taverns.
tower	2	-turm	All references to towers.
weavery	2		All references to wicker/weavery works.
winery	2	Wein, Winzer	No additional notes.
door	1		The only entry is Düren.
dovecote	1		The only entry is Colmar.
gate	1	Tor	The only entry is Solothurn.
glass_works	1	-hütte	All references to glass works.
gravestone	1		The only entry is Krokau.
hall	1	-saal	The only entry is Rauxel.
knights_manor	1		The only entry is Dürrenberg, refering to the manor at the top of the mountain.
mayor_estate	1	mayor	The only entry is Meran.
mine	1		The only entry is Grund (Harz).
post_station	1		The only entry is Neumarkt (Egna), which changed names.
sawmill	1		The only entry is Schneidemühl.
silo	1	-kast	The only entry is Castrop.
stable	1		The only entry is Stuttgart.
table	1	tisk	The only entry is Bitsch (Bitche).
terminal	1		The only entry is Pforzheim.
threshing_floor	1		The only entry is Tennstedt.
CATEGORY: VEGETATION			
tree	147	Birk-, Buch-, Eich-, Erl-, Lind-	Includes any reference to a tree; usually a reference to the type of tree (e.g., Buch, oak); a suffix such as -baum (tree) does not exist; also includes references to their fruit (e.g., apple, nut, pear, plum).
forest	137	-bere/-bire, -forst, -hart, -holz, -lohe, -stock, -wald, -wedel	Includes any reference to forest or woods. -hart is a special case describing a light, moutain forest without having the typical prefix type to the suffix.
clearing	79	-berth/breth, -bracht/-brecht, -grün, -hau/gehäu, -rath/reut/reuth/ried/rode, -scheid, -schneise, -walde	All references to clearings of a forest are considered. -scheid and -schneise indicate border or divides made through clearings, but are not as common as -rode places.
hedge	76	-bram, -dorn, -hag, -hagen, -hain, -heck, -wört, -wurt,	Hedge or other natural fences. Often part of a Landwher, a system of hedges that is either natural, artificial, or both, for protection or for border control/collecting tolls. -wurt is not to be confused with Wurt (artificial mounds to protect against flooding).
bush	17	Dorn-, -busch, -strauch,	All references to bushes, or thorns referring to bushes.
heath	15	-heide	Heath, shrublands, bad soil quality, could also fit in the land category
forest_free	13	-lar	Typically a bald, free spot within a forest.

thicket	10	-ach	A thicket, dense in bushes or trees. Difficult to distinguish with -ach/-bach which means stream.
hurst	9	-horst	No clear translation available except 'hurst'. Typically a bush forest on a slope in a swamp.
garden	7	-garten	No additional notes.
moss	5	Moos-	No additional notes.
swamp_forest	4		No additional notes.
tree_stump	4	-stock, -stumpf	No additional notes.
weathered_plants	4		Fouling, mold, weathered plants
twig	3		No additional notes.
coppice	2		Artificial or managed forest to produce burning wood.
bork	1		The only entry is Lauta (Luty)
bush_forest	1	-hesi/-hais	The only entry is Hesel
mistle	1		The only entry is Mistelbach
		CATEGORY: LAND	
field	108	-feld/felde, -mage, -wang, -wiese	References to fields, often used for agriculture.
swamp	44	-bel/birl/berl, -broich/brook/bruch, -mar/meri	Bruchland are swamps (flooded forested area). Peat does not develop here (in contrast to bogs).
bog	29	-fehn/fen, -moor, -torf	-fehn/fen refers to the bog itself, -torf refers to the peat typically found near bogs.
land	27	-gau, -grün, -kamp, -land/lant, -reich/rich	All references to land are included.
valley	21	-tal/thal	Valley is a popular name after fusions, as they describe the geography of originally separate places. As a rule of thumb, the modern suffix -tal more likely refers to fusions (not counted here) than places with the old spelling thal.
pasture	13	-las, -sweiga, -veluwe/velwa, -weide	All references to pastures, i.e., land designated for grazing.
wallow	11	lache, Lake, liunas, mor, sol, sul	No additional notes.
steep_terrain	10	-bille, -wang	Steep terrain, but a greater emphasis on the land instead of the elevation.
edge_wedge	9	-ecke, -horn, -keil, -winkel	References to the form of an area itself.
bald_spot	7	gola, kahl, kalwa, kalewe	All references to bald spots, whether on land or in mountains.
mud	6	-hor, -mul	No additional notes.
pile	6	-helde/halde	Heap, stockpile.
acre	5	ard/eard/acker, -arten	References to acres, or plowed land.
geest	5	Geest/gest/güst	Dry, infertile land.
ravine	5	-schlucht, -tobel	No additional notes.
swampy_meadow	4	-lug, luh	No additional notes.
wasteland	4	brache, ledo	No additional notes.
cover_guard	3	-hut	E.g., Landeshut and Landshut, meaning "land cover".
fertile_land	3		No additional notes.
remote_enclosed_area	3	-winkel	Reference to remote and enclosed areas around hills, forests, etc.
trough	3		Similar to valley.
area	1		The only entry is Schlangen (long area, snake)
cut	1		A cut in the area
desert	1		The only entry is Wustermark.
ear_form	1	-ohr	The only entry is Oer-Erkenschwick.
elevated_morass	1	-donk, -dunk	The only entry is Wachtendonk.
flat	1	-telva	Flat surface. The only entry is Telfs.
reese	1	rispe	The only entry is Rees.
scrubs_in_swamp	1		The only entry is Labes (Lobez).
straw_land	1		The only entry is Kriens
		CATEGORY: ELEVATION	
mountain	173	-berg, -eck, -gora, -mons, -spitz, -sporn	-berg is typically associated with mountains, -sporn/-spitz describe mountain peaks and -eck refers to mountain ranges. -berg and -eck may also be used as castle names.
elevation	34	barm, Hoch, höhe-, Man-	All references to elevated positions, or places that are located on elevated positions.
stone_cliff	34	-brun, -fels, -klif, -stein	Place on a rock; may sometimes refer to castles (similar to -berg and -stein) so careful etymological evaluation is required.
hill	26	-bühel, -hübel, -hügel	No additional notes.

promontory	10	-horn	A piece of land protruding into the water.
drainage_divide	4	-scheide	Elevated terrain/border separating drainage basins.
foothills	4	-ness, schachen	Transition of plains to elevated positions, i.e., the border to mountains or hills.
ridge	2		All references to ridges.
mudflow	1		The only entry is Rum.
CATEGORY: RESOURCES			
salt	24	-hall, Salz-	-hall is traditionally translated as salt, but could also be translated as 'slope' (e.g., Halle an der Saale).
reed	18	kulba, -ried, schilf	References to reed places as a resource for building properties and to treat water.
gravel	14	Gries, Kies, Sand	All references to gravel or gritted sand.
clay	9	kley, leim, thohe	All references to clay
grain	4	Korn-, Gren-	All references to grain/wheat.
hunting	3		All references to hunting/shooting.
leek_onion	3	Lock/Luck	All references to leek/onions.
honey	2		All references to honey.
iron	2	Eisen, Isern	These references are typically mentioned with -hütt, i.e., referring to iron works, not iron itself.
cabbage	1		The only entry is Kölleda.
chalk	1		The only entry is Kalkar.
flax	1		The only entry is Flachsmeer, first mentioned in 1818, i.e., a modern name.
flint	1		The only entry is Flintbek, in this case considered to refer to a resource and not a stream.
gold	1		The only entry is Goldberg (Zlotoryja), known as Aurum in 1201.
hops	1		The only entry is Hoppegarten, first mentioned in 1797, i.e., a modern name.
peas	1		The only entry is Erwitte.
pitch	1		The only entry Is Schmölln.
rich_in_something	1		Reference to being rich in some resource. The only entry is Bad Reichenall.
slate	1		The only entry is Kirn.
CATEGORY: MISCELLANEOUS			
named_after	42		Places which were named after other places or non-inhabitants. Does not belong to 'no geography' category below, even if the underlying geographic meaning is not clearly identified, as the original place often serves as an inspiration or is modelled after it.
ithi/ede	20	-ithi/ede	There have been debates about the true geographic meaning of -ithi/ede, but no clear consensus has been settled on in the literature.
border_region	10		Place is near some border of the past.
burial	3		All references to burials, including ghosts.
grind	2		All references to grinding, grindstones.
rain	2		Rainy area.
winter	2		No additional notes.
crack	1		The only entry is Mömbris
fog	1		The only entry is Dissen am Teutoburger Wald
holy_cross	1		The only entry is Kreuzlingen
holy_tree	1		The only entry is Altshausen
praying_place	1	Bet-	The only entry is Bedburg
tight	1		The only entry is Spenge
uninhabited	1		The only entry is Eupen in Belgium.
warm	1		The only entry is Tapiau (Gwardeisk).
CATEGORY: NO GEOGRAPHY			
All places without geographic meaning	1,403	any, ballig, bn, borstel/bostel, both, büll, büttel, casa, ec, fleth, gadum/garden, helid/hütte, ica/ici/icz/itz, issa, j/jb/je, jane, k, lage, leben, lev, no, ov/ow, ovici/ovik/owice, plan, rege, sal/saal, salida/selida, sk, teil, trabena, treb, villa, wick/wiek/wick, wies/wis, yna/ynia	**All references without geographic meaning**, typically translated as 'belonging', 'place', inheritance', 'home/house', 'hamlet', 'village', or 'with the people of a person'. Usually the prefix is a personal name (PN) of a male person.
miscellaneous	375		Other references without geographic meanings that do not fall under any of the subcategories below.

-ingen	315	in, ing/ingen/ungen, ingi, inja/unja, inni/innia	All references without geographic meaning based on –ingen. It means “with the people of a PN”.
-heim	256	heim	All references without geographic meaning based on –heim. “Home of PN”.
-dorf	177	dorf/torf	All references without geographic meaning based on –dorf. “Village of PN”.
-hausen	168	haus/hus, hausen/husen	All references without geographic meaning based on –hausen. “House of PN”.
-weiler	52	weier/weil/weiler/wil/wilari/wilen/ville	All references without geographic meaning based on –weiler. “Hamlet of PN”.
-acum	38	-acum/ako, -anum/ianum/um, -orum	All references without geographic meaning based on –acum. Roman suffix.
-beuren	22	beuren/beuern/büren/bur	All references without geographic meaning based on –beuren.

Notes: Frequency is the number of times a characteristic is observed in the dataset. As discussed in the main text, an individual place name may contain several characteristics.

Appendix B

Table B1. Toponyms and urban development: Controlling for time-varying market access (MA)

	(1) Baseline (Table 6, col. 7)		(2) MA before 1200		(3) MA before 1300		(4) MA before 1400		(5) MA before 1500		(6) MA before 1600		(7) MA before 1700		(8) MA before 1800		(9) MA before 1900	
WATER:																		
- Stream	-0.001	(0.01)	-0.002	(0.01)	-0.001	(0.01)	-0.001	(0.01)	-0.001	(0.01)	-0.002	(0.01)	-0.002	(0.01)	-0.002	(0.01)	-0.001	(0.01)
- River	**0.036**	**(0.01)*****	**0.032**	**(0.01)*****	**0.033**	**(0.01)*****	**0.029**	**(0.01)*****	**0.029**	**(0.01)*****	**0.026**	**(0.01)****	**0.026**	**(0.01)****	**0.027**	**(0.01)****	**0.027**	**(0.01)****
- Floodplain	0.012	(0.01)	0.009	(0.01)	0.006	(0.01)	0.003	(0.01)	0.004	(0.01)	0.007	(0.01)	0.008	(0.01)	0.008	(0.01)	0.010	(0.01)
- Ford	**0.066**	**(0.02)*****	**0.071**	**(0.02)*****	**0.068**	**(0.02)*****	**0.064**	**(0.02)*****	**0.061**	**(0.02)*****	**0.050**	**(0.02)****	**0.048**	**(0.02)****	**0.049**	**(0.02)****	**0.044**	**(0.02)***
- Source	-0.034	(0.02)	**-0.037**	**(0.02)***	-0.021	(0.02)	-0.030	(0.02)	**-0.031**	**(0.02)**	**-0.031**	**(0.02)***	**-0.031**	**(0.02)***	**-0.031**	**(0.02)***	**-0.032**	**(0.02)***
- Lake	-0.014	(0.03)	-0.008	(0.03)	-0.010	(0.03)	-0.004	(0.02)	-0.003	(0.02)	-0.002	(0.02)	-0.001	(0.02)	-0.002	(0.02)	-0.003	(0.02)
- Water	0.035	(0.04)	0.037	(0.04)	0.053	(0.04)	**0.061**	**(0.03)***	**0.059**	**(0.03)***	**0.055**	**(0.03)***	0.054	(0.03)	0.052	(0.03)	0.051	(0.03)
- River mouth	0.046	(0.03)	0.040	(0.03)	0.023	(0.03)	0.016	(0.03)	0.008	(0.03)	0.012	(0.03)	0.011	(0.03)	0.011	(0.03)	0.005	(0.03)
- River island	0.022	(0.02)	0.020	(0.02)	0.025	(0.02)	0.010	(0.02)	0.014	(0.02)	0.012	(0.02)	0.014	(0.02)	0.017	(0.02)	0.019	(0.02)
STRUCTURES:																		
- Castle	0.006	(0.01)	0.009	(0.01)	0.002	(0.01)	0.003	(0.01)	0.003	(0.01)	0.002	(0.01)	0.002	(0.01)	0.001	(0.01)	0.001	(0.01)
- Church	**0.061**	**(0.02)*****	**0.056**	**(0.02)*****	**0.052**	**(0.02)*****	**0.046**	**(0.02)*****	**0.045**	**(0.02)*****	**0.044**	**(0.02)*****	**0.044**	**(0.02)*****	**0.046**	**(0.02)*****	**0.048**	**(0.02)*****
- Farm estate	-0.009	(0.02)	-0.007	(0.02)	-0.011	(0.02)	-0.011	(0.02)	-0.010	(0.02)	-0.010	(0.01)	-0.010	(0.01)	-0.011	(0.02)	-0.009	(0.02)
- Monastery	0.020	(0.02)	0.023	(0.02)	0.021	(0.01)	**0.026**	**(0.01)***	0.024	(0.01)	0.023	(0.01)	0.021	(0.01)	0.019	(0.01)	0.022	(0.01)
- Bridge	0.028	(0.03)	0.028	(0.03)	0.018	(0.02)	0.021	(0.02)	0.020	(0.02)	0.011	(0.02)	0.012	(0.02)	0.016	(0.02)	0.013	(0.02)
- Fence	**0.059**	**(0.03)***	**0.075**	**(0.03)****	**0.084**	**(0.03)*****	**0.090**	**(0.03)*****	**0.085**	**(0.03)*****	**0.076**	**(0.03)****	**0.072**	**(0.03)****	**0.076**	**(0.03)****	**0.075**	**(0.03)****
- Mill	0.013	(0.04)	0.011	(0.04)	0.005	(0.04)	0.005	(0.03)	0.003	(0.03)	0.002	(0.03)	0.003	(0.03)	0.003	(0.03)	0.005	(0.03)
VEGETATION:																		
- Tree	0.022	(0.01)	0.020	(0.01)	0.021	(0.01)	0.021	(0.01)	0.022	(0.01)	0.020	(0.01)	0.019	(0.01)	0.019	(0.01)	0.019	(0.01)
- Forest	-0.029	(0.05)	-0.028	(0.05)	-0.020	(0.05)	-0.020	(0.05)	-0.021	(0.05)	-0.024	(0.05)	-0.023	(0.05)	-0.021	(0.05)	-0.021	(0.05)
- Clearing	**0.054**	**(0.02)*****	**0.051**	**(0.02)*****	**0.056**	**(0.02)*****	**0.050**	**(0.02)*****	**0.050**	**(0.02)*****	**0.051**	**(0.02)*****	**0.050**	**(0.02)*****	**0.049**	**(0.02)*****	**0.051**	**(0.02)*****
- Hedge	**-0.069**	**(0.02)*****	**-0.064**	**(0.02)*****	**-0.051**	**(0.02)*****	**-0.052**	**(0.02)*****	**-0.050**	**(0.02)*****	**-0.044**	**(0.02)*****	**-0.045**	**(0.02)*****	**-0.046**	**(0.02)*****	**-0.042**	**(0.02)*****
- Bush	0.010	(0.03)	0.006	(0.03)	-0.012	(0.03)	-0.002	(0.03)	-0.006	(0.03)	-0.012	(0.03)	-0.007	(0.03)	-0.006	(0.03)	-0.004	(0.03)
- Heath	-0.431	(0.44)	-0.446	(0.44)	-0.433	(0.44)	-0.433	(0.44)	-0.429	(0.44)	-0.432	(0.44)	-0.435	(0.44)	-0.437	(0.44)	-0.438	(0.44)
LAND:																		
- Field	-0.009	(0.02)	-0.010	(0.02)	-0.006	(0.01)	-0.005	(0.01)	-0.005	(0.01)	-0.001	(0.01)	-0.002	(0.01)	-0.003	(0.01)	-0.000	(0.01)
- Swamp	-0.018	(0.02)	-0.006	(0.02)	-0.005	(0.02)	-0.017	(0.02)	-0.017	(0.02)	-0.021	(0.02)	-0.015	(0.02)	-0.014	(0.02)	-0.014	(0.02)
- Bog	0.007	(0.03)	-0.003	(0.03)	**-0.055**	**(0.03)***	-0.036	(0.03)	-0.034	(0.03)	-0.022	(0.03)	-0.025	(0.03)	-0.028	(0.03)	-0.025	(0.03)
- Land	0.021	(0.03)	0.007	(0.03)	0.012	(0.03)	-0.010	(0.03)	-0.006	(0.03)	-0.013	(0.03)	-0.013	(0.03)	-0.010	(0.03)	-0.016	(0.03)
- Valley	0.015	(0.04)	0.013	(0.04)	0.007	(0.03)	0.003	(0.03)	0.002	(0.03)	0.003	(0.03)	0.004	(0.03)	0.005	(0.03)	0.008	(0.03)
ELEVATION:																		
- Mountain	0.014	(0.01)	0.010	(0.01)	0.014	(0.01)	0.013	(0.01)	0.012	(0.01)	0.009	(0.01)	0.008	(0.01)	0.010	(0.01)	0.012	(0.01)
- Elevation	0.006	(0.02)	0.007	(0.02)	0.001	(0.02)	0.009	(0.02)	0.008	(0.02)	0.012	(0.02)	0.011	(0.02)	0.010	(0.02)	0.011	(0.02)
- Stone/cliff	0.005	(0.02)	0.008	(0.02)	0.017	(0.02)	0.016	(0.02)	0.016	(0.02)	0.016	(0.02)	0.014	(0.02)	0.013	(0.02)	0.012	(0.02)
- Hill	**0.041**	**(0.02)***	0.027	(0.02)	**0.034**	**(0.02)***	**0.033**	**(0.02)***	0.030	(0.02)	0.030	(0.02)	0.030	(0.02)	**0.031**	**(0.02)***	0.028	(0.02)
RESOURCES:																		
- Salt	0.014	(0.03)	-0.001	(0.03)	-0.001	(0.03)	0.003	(0.02)	0.000	(0.02)	-0.002	(0.02)	-0.002	(0.02)	-0.002	(0.02)	-0.007	(0.02)
- Reed	0.024	(0.02)	0.024	(0.02)	**0.035**	**(0.02)***	0.024	(0.02)	0.026	(0.02)	0.026	(0.02)	0.026	(0.02)	0.026	(0.02)	0.023	(0.02)
CONTROLS:																		
No geography	**-0.015**	**(0.01)***	**-0.014**	**(0.01)***	-0.010	(0.01)	-0.012	(0.01)	-0.012	(0.01)	-0.011	(0.01)	-0.011	(0.01)	**-0.012**	**(0.01)***	-0.011	(0.01)
log(UP)	**0.948**	**(0.03)*****	**0.950**	**(0.03)*****	**0.952**	**(0.03)*****	**0.953**	**(0.03)*****	**0.953**	**(0.03)*****	**0.954**	**(0.03)*****	**0.954**	**(0.03)*****	**0.954**	**(0.03)*****	**0.956**	**(0.03)*****
First record	**-0.073**	**(0.04)***	**-0.080**	**(0.04)****	**-0.076**	**(0.04)***	**-0.070**	**(0.04)***	**-0.067**	**(0.04)***	**-0.068**	**(0.04)***	**-0.068**	**(0.04)***	**-0.067**	**(0.04)***	**-0.066**	**(0.04)***
City rights	**0.130**	**(0.03)*****	**0.126**	**(0.03)*****	**0.127**	**(0.03)*****	**0.126**	**(0.03)*****	**0.126**	**(0.03)*****	**0.126**	**(0.03)*****	**0.125**	**(0.03)*****	**0.125**	**(0.03)*****	**0.122**	**(0.03)*****
log(MA)																		
- 0 to 20km			**-0.008**	**(0.00)*****	**-0.018**	**(0.00)*****	**-0.015**	**(0.00)*****	**-0.020**	**(0.00)*****	**-0.015**	**(0.00)*****	**-0.014**	**(0.00)*****	**-0.014**	**(0.00)*****	**-0.015**	**(0.00)*****
- 20 to 50km			**-0.012**	**(0.01)****	**-0.033**	**(0.01)*****	**-0.037**	**(0.01)*****	**-0.044**	**(0.01)*****	**-0.090**	**(0.01)*****	**-0.097**	**(0.01)*****	**-0.094**	**(0.01)*****	**-0.109**	**(0.01)*****
- 50 to 100km			**-0.041**	**(0.00)*****	**-0.048**	**(0.01)*****	**-0.115**	**(0.02)*****	**-0.097**	**(0.02)*****	**-0.077**	**(0.02)*****	**-0.062**	**(0.02)*****	**-0.054**	**(0.02)*****	**-0.046**	**(0.02)*****
Country FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lat & Lon	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Observations	3,705		3,705		3,705		3,705		3,705		3,705		3,705		3,705		3,705	
Adjusted R^2	0.969		0.970		0.971		0.971		0.971		0.972		0.972		0.972		0.972	

Notes: Dependent variable is log(1910 population). Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Column (1) reports the baseline for the full sample. Columns (2) through (9) include log(market access) based on the specified period and 0-20km, 20-50km, and 50-100km distance bands. Parameter estimates for the Miscellaneous category, characteristics with <15 observations, and the constant are not displayed for brevity.

Table B2. Geography and early geographical advantage: Controlling for time-varying market access (MA)

	(1) Baseline (Table 8, col. 1)		(2) MA before 1200		(3) MA before 1300		(4) MA before 1400		(5) MA before 1500		(6) MA before 1600		(7) MA before 1700		(8) MA before 1800		(9) MA before 1900	
WATER:																		
- Stream	**0.217**	**(0.09)****	**0.228**	**(0.09)*****	**0.212**	**(0.09)****	**0.204**	**(0.09)****	**0.201**	**(0.09)****	**0.199**	**(0.09)****	**0.201**	**(0.09)****	**0.198**	**(0.09)****	**0.203**	**(0.09)****
- River	-0.063	(0.08)	-0.059	(0.08)	-0.070	(0.08)	-0.070	(0.08)	-0.074	(0.08)	-0.075	(0.08)	-0.074	(0.08)	-0.077	(0.08)	-0.071	(0.08)
- Floodplain	-0.178	(0.17)	-0.163	(0.17)	-0.189	(0.17)	-0.188	(0.17)	-0.189	(0.17)	-0.191	(0.17)	-0.190	(0.17)	-0.195	(0.17)	-0.184	(0.17)
- Ford	-0.080	(0.21)	-0.073	(0.20)	-0.088	(0.21)	-0.092	(0.21)	-0.094	(0.21)	-0.097	(0.21)	-0.096	(0.21)	-0.097	(0.21)	-0.087	(0.21)
- Source	**0.360**	**(0.22)***	**0.376**	**(0.21)***	**0.375**	**(0.22)***	**0.362**	**(0.21)***	**0.359**	**(0.21)***	**0.360**	**(0.21)***	**0.361**	**(0.21)***	**0.368**	**(0.21)***	**0.356**	**(0.21)***
- Lake	-0.326	(0.21)	-0.313	(0.21)	-0.311	(0.21)	-0.315	(0.21)	-0.327	(0.21)	-0.330	(0.21)	-0.329	(0.21)	-0.330	(0.21)	-0.333	(0.22)
- Water	-0.126	(0.26)	-0.109	(0.26)	-0.075	(0.26)	-0.089	(0.26)	-0.100	(0.26)	-0.100	(0.26)	-0.100	(0.26)	-0.095	(0.26)	-0.093	(0.26)
- River mouth	-0.163	(0.30)	-0.180	(0.30)	-0.182	(0.30)	-0.168	(0.31)	-0.171	(0.31)	-0.163	(0.31)	-0.163	(0.31)	-0.164	(0.31)	-0.149	(0.31)
- River island	**0.399**	**(0.21)***	**0.408**	**(0.22)***	**0.423**	**(0.20)****	**0.402**	**(0.21)***	**0.399**	**(0.21)***	**0.402**	**(0.20)****	**0.402**	**(0.20)****	**0.403**	**(0.20)****	**0.400**	**(0.20)****
STRUCTURES:																		
- Castle	**-0.330**	**(0.09)*****	**-0.332**	**(0.09)*****	**-0.334**	**(0.09)*****	**-0.332**	**(0.09)*****	**-0.331**	**(0.09)*****	**-0.332**	**(0.09)*****	**-0.333**	**(0.09)*****	**-0.333**	**(0.09)*****	**-0.325**	**(0.09)*****
- Church	**0.257**	**(0.14)***	**0.269**	**(0.14)***	**0.258**	**(0.14)***	**0.249**	**(0.14)***	**0.250**	**(0.14)***	**0.251**	**(0.14)***	**0.250**	**(0.14)***	**0.247**	**(0.14)***	**0.245**	**(0.14)***
- Farm estate	**0.407**	**(0.13)*****	**0.397**	**(0.13)*****	**0.393**	**(0.13)*****	**0.395**	**(0.13)*****	**0.393**	**(0.13)*****	**0.392**	**(0.13)*****	**0.393**	**(0.13)*****	**0.389**	**(0.13)*****	**0.393**	**(0.13)*****
- Monastery	0.031	(0.22)	0.016	(0.22)	0.025	(0.22)	0.034	(0.22)	0.035	(0.22)	0.035	(0.22)	0.035	(0.22)	0.034	(0.22)	0.048	(0.22)
- Bridge	**-0.440**	**(0.25)***	**-0.430**	**(0.25)***	**-0.441**	**(0.25)***	**-0.448**	**(0.25)***	**-0.455**	**(0.26)***	**-0.459**	**(0.26)***	**-0.454**	**(0.26)***	**-0.452**	**(0.26)***	**-0.429**	**(0.26)***
- Fence	0.152	(0.20)	0.141	(0.20)	0.185	(0.21)	0.173	(0.21)	0.170	(0.21)	0.171	(0.21)	0.170	(0.21)	0.173	(0.20)	0.167	(0.20)
- Mill	-0.232	(0.26)	-0.231	(0.26)	-0.243	(0.26)	-0.240	(0.26)	-0.242	(0.26)	-0.240	(0.26)	-0.238	(0.26)	-0.242	(0.26)	-0.252	(0.26)
VEGETATION:																		
- Tree	0.086	(0.11)	0.089	(0.11)	0.090	(0.11)	0.084	(0.11)	0.083	(0.11)	0.081	(0.11)	0.083	(0.11)	0.082	(0.11)	0.088	(0.12)
- Forest	0.188	(0.12)	0.191	(0.12)	0.193	(0.12)	0.186	(0.12)	0.184	(0.12)	0.183	(0.12)	0.186	(0.12)	0.187	(0.12)	0.185	(0.12)
- Clearing	0.209	(0.19)	0.222	(0.19)	0.221	(0.19)	0.226	(0.19)	0.226	(0.19)	0.224	(0.19)	0.223	(0.19)	0.221	(0.19)	0.223	(0.19)
- Hedge	-0.149	(0.30)	-0.155	(0.30)	-0.118	(0.30)	-0.130	(0.30)	-0.128	(0.29)	-0.123	(0.29)	-0.127	(0.30)	-0.124	(0.30)	-0.141	(0.29)
- Bush	0.004	(0.47)	0.027	(0.47)	0.012	(0.47)	0.037	(0.46)	0.037	(0.46)	0.034	(0.46)	0.023	(0.47)	0.025	(0.47)	0.026	(0.47)
- Heath	-0.114	(0.40)	-0.134	(0.39)	-0.090	(0.40)	-0.122	(0.40)	-0.122	(0.40)	-0.120	(0.40)	-0.121	(0.40)	-0.122	(0.40)	-0.117	(0.39)
LAND:																		
- Field	0.067	(0.12)	0.085	(0.12)	0.077	(0.12)	0.079	(0.12)	0.073	(0.12)	0.072	(0.12)	0.070	(0.12)	0.069	(0.12)	0.066	(0.12)
- Swamp	**0.299**	**(0.17)***	**0.292**	**(0.17)***	**0.295**	**(0.17)***	**0.295**	**(0.16)***	**0.293**	**(0.16)***	**0.292**	**(0.16)***	**0.303**	**(0.16)***	**0.307**	**(0.16)***	**0.311**	**(0.16)***
- Bog	0.358	(0.44)	0.352	(0.45)	0.300	(0.46)	0.303	(0.45)	0.310	(0.45)	0.283	(0.46)	0.286	(0.46)	0.272	(0.46)	0.290	(0.46)
- Land	-0.060	(0.28)	-0.041	(0.28)	-0.078	(0.28)	-0.076	(0.28)	-0.079	(0.28)	-0.082	(0.28)	-0.075	(0.28)	-0.079	(0.28)	-0.086	(0.28)
- Valley	0.259	(0.38)	0.269	(0.37)	0.250	(0.37)	0.243	(0.37)	0.244	(0.37)	0.236	(0.37)	0.239	(0.37)	0.239	(0.37)	0.246	(0.36)
ELEVATION:																		
- Mountain	-0.187	(0.14)	-0.180	(0.14)	-0.187	(0.14)	-0.191	(0.14)	-0.191	(0.14)	-0.194	(0.14)	-0.193	(0.14)	-0.194	(0.14)	-0.194	(0.14)
- Elevation	0.076	(0.18)	0.086	(0.18)	0.098	(0.18)	0.088	(0.18)	0.090	(0.18)	0.091	(0.18)	0.088	(0.18)	0.084	(0.18)	0.073	(0.17)
- Stone/cliff	-0.269	(0.28)	-0.259	(0.28)	-0.258	(0.27)	-0.275	(0.28)	-0.279	(0.28)	-0.280	(0.28)	-0.278	(0.28)	-0.280	(0.28)	-0.278	(0.28)
- Hill	0.433	(0.41)	0.441	(0.42)	0.435	(0.41)	0.448	(0.40)	0.447	(0.40)	0.451	(0.40)	0.449	(0.40)	0.444	(0.40)	0.460	(0.40)
RESOURCES:																		
- Salt	0.061	(0.21)	0.055	(0.22)	0.036	(0.21)	0.042	(0.21)	0.038	(0.21)	0.037	(0.21)	0.041	(0.21)	0.038	(0.21)	0.046	(0.21)
- Reed	-0.202	(0.29)	-0.200	(0.29)	-0.194	(0.28)	-0.196	(0.29)	-0.193	(0.29)	-0.190	(0.29)	-0.191	(0.29)	-0.181	(0.29)	-0.190	(0.29)
CONTROLS:																		
No geography log(MA)	**0.163**	**(0.07)****	**0.165**	**(0.07)****	**0.161**	**(0.07)****	**0.158**	**(0.07)****	**0.156**	**(0.07)****	**0.155**	**(0.07)****	**0.155**	**(0.07)****	**0.155**	**(0.07)****	**0.155**	**(0.07)****
- 0 to 20km			**0.058**	**(0.02)*****	0.027	(0.02)	**0.048**	**(0.03)***	**0.056**	**(0.03)***	**0.062**	**(0.03)****	**0.056**	**(0.03)***	**0.060**	**(0.03)****	**0.079**	**(0.03)****
- 20 to 50km			-0.016	(0.02)	-0.033	(0.04)	-0.018	(0.06)	-0.036	(0.06)	-0.060	(0.06)	-0.050	(0.06)	-0.051	(0.06)	-0.016	(0.07)
- 50 to 100km			0.021	(0.04)	**-0.182**	**(0.08)****	**-0.211**	**(0.11)****	**-0.200**	**(0.11)***	**-0.201**	**(0.11)***	**-0.195**	**(0.11)***	**-0.229**	**(0.11)****	**-0.257**	**(0.12)****
Country FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Lat & Lon	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Bicentennial FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Observations	2,004		2,004		2,004		2,004		2,004		2,004		2,004		2,004		2,00	
Adjusted R^2	0.346		0.347		0.348		0.347		0.347		0.348		0.347		0.348		0.349	

Notes: Dependent variable is the log of the difference between the year that a place received city rights and its first record. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Column (1) reports the baseline for the full sample. Columns (2) through (9) include log(market access) based on the specified period and 0-20km, 20-50km, and 50-100km distance bands. Parameter estimates for the Miscellaneous category, characteristics with <15 observations, controls and the constant are not displayed for brevity.