Journal of Urban Economics 98 (2017) 139–157

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

City seeds: Geography and the origins of the European city system

Maarten Bosker[a,b,*], Eltjo Buringh[b]

[a] Erasmus University of Rotterdam, CEPR, The Tinbergen Institute, The Netherlands
[b] Utrecht University, The Netherlands

ARTICLE INFO

Article history:
Received 20 December 2013
Revised 23 August 2015
Available online 11 September 2015

JEL classification:
N93
O18
R10

Keywords:
City origins
Economic geography
Europe

ABSTRACT

Cities are the focal points of the world economy. This paper sheds new empirical light on their origins. Using a new dataset covering over 250,000 randomly selected potential city locations, and all actual cities during the period 800–1800, we disentangle the different roles of geography in shaping today's European city system. We find that a location's *physical geography* characteristics are the dominant determinants of its urban chances. Preferential location for water- or land-based transportation is a particularly important city seed. In addition, a location's *position relative* to already-existing cities matters for its urban chances. Interestingly, it does so in a way corresponding to predictions from economic geography theory.

"*In a more advanced era, when better methods would permit man to conquer Nature […], it would doubtless have been possible to build towns anywhere the spirit of enterprise and the quest of gain might suggest a site. But it was quite another matter in a period when society had not yet acquired enough vigor to rise above the physical conditions in the midst of which it developed. […] the towns of the Middle Ages were a phenomenon determined as much by physical surroundings as the course of rivers is determined by the conformation of the mountains and the direction of the valleys.*"

Henri Pirenne, 1925, p. 138/39

1. Introduction

Today the European landscape is dotted with cities. Historically this was quite different. In the early medieval period Europe only knew a handful of cities. Over the next millennium this changed dramatically. Cities started to appear on an unprecedented scale, and virtually everywhere on the continent. Fig. 1 shows that in 800, we only find a few scattered cities in mainly Spain, Italy, France and Germany. One thousand years later, cities have appeared all over the continent.[1]

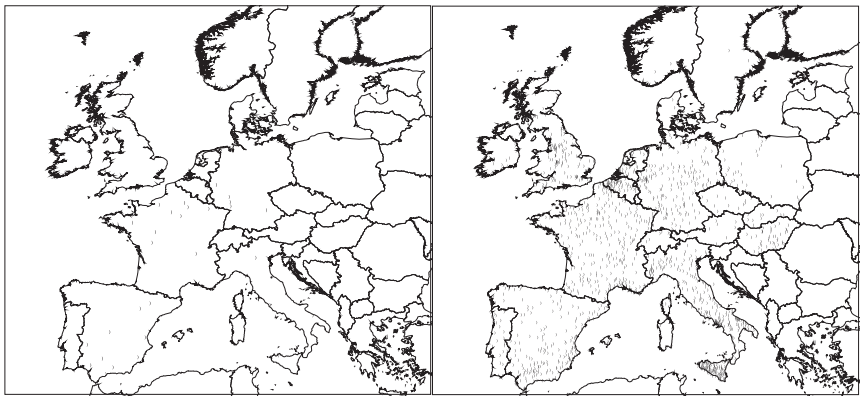
The rise of the city in the European landscape is important for several reasons. Throughout history, cities have been the important loci for technological innovation, institutional progress, (international) trade, political power, and culture (Pirenne, 1925; Glaeser, 2011). Also, cities are generally more productive places. The concentration of many people e.g. allows for a greater degree of specialization, carries positive externalities such as knowledge spillovers, and facilitates a more efficient provision of public goods (Lampard, 1955; Marshall, 1890). It may therefor not be surprising that cities are argued to have played a very important role in Europe's economic 'take-off' during the late Medieval and Early Modern period. Urbanization and economic development often go hand in hand (Acemoglu et al., 2005; De Vries, 1984; Galor, 2005).

The importance of cities in the development process makes understanding their origins of great interest. Cities do not develop everywhere. The question '*why do cities form in some locations, and not, or only much later, in others?*' lies at the heart of this paper. In

[1] Fig. A1 in Appendix A further illustrates the rise of the city in the European landscape. Over our sample period, Europe's urbanization rate increased from only 3% in 800 to 15% in 1800. Urban population increased 30-fold from 0.7 to 21 million, whereas total population increased 6-fold from 23 to 137 million. A full, century-by-century, visualization of the formation of the European city system is available upon request.

* Corresponding author at: P.O. Box 1738, 3000DR Rotterdam, The Netherlands.
E-mail address: bosker@ese.eur.nl (M. Bosker).

http://dx.doi.org/10.1016/j.jue.2015.09.003
0094-1190/



(a) 800 (b) 1800

Fig. 1. The European city system in 800 and 1800. *Notes*: cities are denoted by black dots [see Section 3.2 for more detail on the city definition used]. In 800 there are 34 cities, in 1800 this number has increased to over 1450.

particular, we empirically uncover the role(s) of geography, widely viewed as the most important determinant of a location's urban chances, in 'sowing the seeds' of the European city system.

The narrative urban history, the economic geography, as well as the more recent urban economic and new economic geography literatures, stress two important, but very different, roles for geography in the origins of an urban system.[2]

First, geography determines a location's physical, or *1st nature geography*, characteristics. A location's agricultural potential, access to natural resources, its transportation possibilities and its defensive advantages, have all been noted as important city seeds. The second role for geography, although already stressed by e.g. von Thünen (1826), Christaller (1935) and Lösch (1938, 1940), has received renewed attention in the economics literature following Krugman (1991, 1993b). While not denying the importance of 1st nature geography, this line of literature stresses the importance of a location's position relative to the already existing urban system, its *2nd nature geography*, for its urban prospects. Locations far away from already existing cities, although maybe well suited for urban development based on their own characteristics, do not benefit from the ease of access to the existing cities' markets, and as a result "*remained sterile, like seed fallen upon stony ground*" (Pirenne, 1925, p. 145). On the other hand, already existing cities may also pose strong limits on urban development in their (immediate) backyard, reducing the urban chances of nearby locations.

The debate on the relevance of the two different roles of geography in determining cities' origins has up to now largely taken place without using rigorous empirical evidence. Instead, it relies on historical narratives, largely descriptive accounts of European urbanization, or detailed case studies looking at one particular city or region only. Some empirical papers do look at the relative importance of 1st and 2nd nature geography, but these papers typically focus on the evolution of a city after its initial establishment. They e.g. look for path-dependence in urban development, or identify the important determinants of a city's size (see Bleakley and Lin, 2010; Davis and Weinstein, 2002; Bosker et al., 2013; or Redding and Sturm, 2008). By looking at city size *conditional* on a city's existence, although very interesting in itself, these papers effectively take cities' location as given and refrain from shedding empirical light on the question why these cities formed at their particular locations in the first place. They do not answer the question why other, often a priori equally viable, locations never became a city or only did so at a much later stage.

The aim of our paper is to identify the important determinants of city location. It allows us to shed light on predictions from (economic) theory that are specific to city emergence.

We focus on the emergence of the European city system in particular.[3] The European case provides an ideal testing ground. First, historical data availability on the size and characteristics of individual cities in Europe is better than that of other continents in terms of both spatial and temporal coverage. This is largely due to the work of Bairoch et al. (1988) and De Vries (1984). They have constructed comprehensive data sets providing population estimates for many cities in Europe starting as early as the year 800. We build on this data in two ways. First, we extend its coverage by also considering over 250,000 randomly selected potential city locations: locations that in principle could have become a city but never did. Second, we complement the existing datasets with detailed information on each location's 1st and 2nd nature geography characteristics.

All this data is available for the period, 800–1800, during which one can forcefully argue that the seeds for the eventual European city system were sown.[4] Following the eclipse of the Roman empire, cities in Europe withered (Pirenne, 1925; Greif, 1992). In 800, only about 30 cities can be found. Over the next millennium Europe witnessed an unprecedented revival of urban activity and the establishment of cities on a scale not seen before (Davis, 1955, p. 432). In 1800 their number had increased to over 1450 (see Fig. 1), effectively laying out the foundations of today's European city system.

[2] Influential contributions in these literatures are e.g. Pirenne (1925), De Vries (1984) or Bairoch (1988) [urban history], Christaller (1935), Lösch (1940), Ullman (1941) or Lampard (1955) [economic geography], and Krugman (1993a), Fujita and Mori (1997) or Behrens (2007) [urban economics/new economic geography].

[3] We define Europe as roughly everything west of the line Trieste – St. Petersburg (with the exception of the Baltic States). This line is based on the European Marriage Pattern (see Hajnal, 1965), and coincides with the border of the Catholic Church during the Middle Ages. See also De Vries (1984) or Findlay and O'Rourke (2007). Europe thus defined comprises current-day Norway, Sweden, Finland, Poland, Germany, Czech Republic, Slovakia, Austria, Hungary, Belgium, Luxembourg, Netherlands, France, Great Britain, Ireland, Switzerland, Italy, Spain and Portugal.

[4] We focus on the 800–1800 period for the following reasons. We start in 800 as it is the first year for which comprehensive data on city population exists for Europe, i.e. Bairoch et al. (1988). We stop in 1800 because not doing so would add the Industrial Revolution to our sample (see e.g. Ashton, 1948). The substantial changes during that period in terms of transportation (railroads, steamships), production (both industrial and agricultural), and the importance of different natural resources (coal), turned many locations that previously had little chance of becoming a city into viable city sites (e.g. many locations in the coal-rich areas of Germany, Sweden, north-east England, and the Limburg provinces of both Belgium and The Netherlands). Including the Industrial Revolution in our view requires a detailed account of its effects, which lies beyond the scope of this paper.

Using our data set, we quantify the different roles of geography in conditioning the spread of cities across the European continent. We explicitly base our empirical analysis on the main theoretical insights regarding the role of 1st and 2nd nature geography in sowing the seeds of cities. These insights come from both the economic and urban history literature, as well as from the more recent new economic geography literature. They serve as the theoretical underpinnings of our empirical analysis, guiding the choice of 1st and 2nd nature geography variables to include in our empirical analysis.

We find that both 1st and 2nd nature geography played an important role in the origins of the European city system. First nature geography is the dominant determinant of a location's urban chances throughout the 800–1800 period that we study. Preferential location for water- or land-based transportation is a particularly important city seed. Second nature geography also matters, but to lesser extent. Interestingly, the effect that an already existing city exerts on the urban chances of its surroundings, as well as the way it changes over the centuries, corresponds closely to predictions made by (new) economic geography theory.

2. Insights from theory

2.1. Economic and urban history

Traditionally, the debate on cities' origins was conducted within the realm of the, largely narrative, economic and urban history literature (see e.g. Pirenne, 1925; Weber, 1922; Bairoch, 1988; De Vries, 1984). This literature stresses a priori differences between locations as the main reason why some locations are more likely than others to develop into a city. Such spatial inhomogeneities, what we call *1st nature geography*, arise most notably from differences between locations in their resource abundance or transportation possibilities.

Attractive city locations were those close to natural resources (fertile plains, mineral deposits, thermal springs, etc.) and locations with good access to the main trade routes. Given that the city relies on exchange with its hinterland, location on a navigable river, an overland transport route, or at sea offers substantial advantages in terms of transportation possibilities (Bleakley and Lin (2010) aptly illustrate this for portage sites in the US).

A location's defensive opportunities and its climatic conditions are also often mentioned (Pirenne, 1925, pp. 72–76; Hohenberg and Lees, 1995, p. 30). A favorable climate for agricultural production, or a strategic location at a mountain pass or hill overlooking the countryside, makes locations more attractive city sites. A location's transportation possibilities however, are mostly viewed to overshadow these other motives. As put by Bairoch (1988, p. 143) "*The critical role played by transport in the location of cities does not rule out exceptions, but statistically speaking these are in the minority*."

2.2. Economic geography

Spatial inhomogeneities also feature prominently in the economic (geography) literature on city creation (Anas et al., 1998; Fujita and Mori, 1996; Krugman, 1993a; Behrens, 2007; Konishi, 2000). Although this literature does not deny that soil quality or climate are important determinants of city location, it views trade costs as one of the most crucial elements in the process of city formation.[5]

They are vital to a city given that it relies on exchange with its hinterland to meet its own demand for agricultural produce. When the cost of transporting agricultural goods (or the goods the city produces in return) are very high, this results in the so-called tyranny of distance, and cities only form in locations where sufficient food can be imported from nearby (see Duranton, 1999, p. 2173). However, when trade costs diminish due to e.g. improvements in transportation technology or lower trade barriers (lower tolls, safer roads, improved freight insurance, etc.), the tyranny of distance is alleviated and the (relative) importance of 1st nature geography diminishes. Since agricultural products can now be shipped over longer distances at lower costs, it becomes possible to establish cities at locations that, given their lack of 1st nature geography advantages, were previously unviable to host a city.

Still, even with a diminishing importance of 1st nature geography, not all locations become equally viable future city sites. This crucially depends on their *2nd nature geography characteristics*, i.e. their position relative to the urban system already in place. Earlier contributions by e.g. Christaller (1935), Lösch (1938, 1940), or Ullman (1941) already stressed that "*no city is ever an island existing in and of itself*" (Lampard, 1955). Yet, it was only fairly recent that several papers explicitly focus on the *where-do-cities-form* question in a theoretical framework of endogenous city location. These papers (e.g. Fujita and Mori, 1996, 1997; Fujita et al., 1999; Behrens, 2007) not only establish theoretically, using fully specified general equilibrium models, under what conditions a new city will form, they also make clear predictions about *where these new cities will most likely form*. Other urban economic theories typically remain silent on this *where do cities form*-question. A city's relative location either bears no consequences for its own development (e.g. Henderson, 1974; Black and Henderson, 1999), or the number of cities is fixed so that no new cities emerge (e.g. Eeckhout, 2004; Duranton, 2007; but also classic contributions by von Thünen (1826), Christaller (1935) or Lösch (1940)).

In this paper, we take the most important prediction of these models of endogenous city location to the data. In particular (see the Theory Appendix for a more detailed exposition of one of these models), they predict that a location's urban chances depend first and foremost on *its distance to an already existing urban center*. Locations too close to an already existing city face too strong competition with that city, both for agricultural produce and for inhabitants.[6] On the other hand, locations too far from an already existing city cannot take full advantage of the trading possibilities with the already existing city. This leaves *locations at medium range from existing cities as preferred new city locations*: they offer relatively cheap trading possibilities with the already existing cities compared to locations further off, as well as only limited competition with these same existing cities compared to locations at closer range. It is this non-linear effect exerted by an already existing city on the urban chances of its surroundings that we take to the data in our empirical sections.

In doing so, we however face two important difficulties that both relate to establishing the importance of 2nd nature geography. First and foremost, it is very difficult to disentangle the effect of 2nd nature geography from unobserved 1st nature geography characteristics that are clustered in space. We discuss this very important concern, and the different ways in which we alleviate it as best as we can in detail throughout the paper. Second, in all the above-mentioned economic geography models, the spatial extent of the positive effect of an existing city at medium range on another location's urban chances, typically depends delicately on the magnitude of trade costs, the extent of agglomeration economies

[5] Besides trade costs, the extent of agglomeration economies in urban production, and the costs associated with (economic) density play an important role in these models. Trade costs are all costs associated with moving goods from one location to another, including not only transportation costs but also tolls, tariffs and less tangible costs associated with differences in e.g. language, institutions or culture.

[6] Not uncommon in medieval times, the existing city may even use force to prevent a competitor city forming in its immediate backyard. Or, less violently, put severe restrictions on any economic activity in its immediate vicinity. The German 'Bannmeile' is a good example (see Ennen, 1972).

in urban production, congestion costs, and overall population size[7] (see e.g. Eq. (B2) and Fig. B1 in Appendix B). Unfortunately, data on most of these variables is unavailable during the 800–1800 period[8]; one exception being overall population numbers. As a result, we can verify whether or not the patterns in the data are consistent with theory, but cannot unambiguously ascribe them to (developments in) one or more of the important factors stressed by theory as explaining these patterns.

3. Data and descriptives

We focus in turn on our choice of potential city locations, the city-definition that we employ, and on the 1st and 2nd nature geography variables we collected.[9]

3.1. Potential city locations

In order to empirically study the rise of cities in Europe, the first choice to make is which locations to consider as potential city locations. We decided not to a priori exclude any location, and in principle consider all possible locations (on land) as potential city locations. To do our empirics, we therefor draw a (very) large random sample from the entire population of possible city locations, and consider 259,776 randomly drawn coordinate pairs as our baseline sample. Note that one could also interpret our sampling strategy as randomly sampling individual grid cells from a very fine exhaustive grid overlaying the countries in our sample.[10] These countries cover an area of about 3.5 million km^2, so that this grid consists of roughly 350 million grid cells of about 100 m × 100 m each [see Footnotes 10 and 11 for more detail on our choice of empirical design].[11]

Fig. 2 shows our sample. Given the large number of randomly drawn locations, plotting all of them (Fig. 2a) basically reproduces the entire map of Europe. It clearly shows the density as well as the boundaries of our sample.[12]

Fig. 2b, zooming in on the English Channel, shows that Fig. 2a is indeed made up of the above-mentioned 259,776 randomly drawn coordinate pairs. To put some numbers on the spatial detail of our sample; the median [average] distance between a randomly drawn coordinate pairs and its closest neighbor is only 1.67 km [1.79 km]. It is important to note that our sampling design does result in an oversampling of locations at higher latitudes (i.e. it does not explicitly take the Earth's curvature into account). This is especially pregnant for the Nordic countries and Scotland. This oversampling, while in principle problematic, does in practice not crucially affect our inference. Hardly any cities can be found in these areas over our sample period. They provide little useful variation to identify the effects we are after in this paper, especially when identifying off of variation at a geographically localized scale (e.g. when including NUTS2/century fixed effects, see the next section for more on this).[13]

3.2. Actual cities – definition and location

Next, we need information on where, and when, actual cities appear. This requires us to define what we mean by a city. In all our baseline analyses, we define a city as an agglomeration of at least 5000 inhabitants. In doing so, we basically adopt the definition proposed by both De Vries (1984) and Bairoch (1988). Of course, this absolute size criterion of 5000 inhabitants may in certain cases be too low and thus wrongly ascribe an urban role to a location (see e.g. Malanima (1998) on Sicilian agrotowns). On the other hand, the opposite, i.e. the cutoff being too high, has also been argued, especially for the early medieval period (see e.g. Dyer, 1995). Both Bairoch (1988, pp. 137/138) and De Vries (1984, pp. 53/54 or 21/22) certainly do not deny this but both view the use of a population cutoff of 5000 inhabitants as providing "*a criterion that may be questionable in certain respects but which nevertheless remains for all that the most adequate and especially the most operational*." (Bairoch, 1988, p. 494).[14]

The alternative would be to define cities on the basis of more criteria than population size only (e.g. having city rights or certain economic, religious or institutional features). In the words of Bairoch (1988) this would however be "*much less operational*" (p. 494). Not only would it constitute a very time consuming exercise; to agree on what features a location needs to have in order to qualify as a city would be subject to much debate. Are city rights sufficient, or should it also have a fair, a market and/or a mint to qualify as a city? And, if so, should these fairs or markets be of a certain size, or of regional importance, before a location qualifies as a city? Even if we were to agree on which features to include in this city definition (and data on all these features would be readily available), the substantial institutional, political and religious differences between the different societies in Europe further complicates the task of consistently applying this definition

[7] For particular configurations of these model parameters, it can even be the case that only one city, or even no city, emerges (see Appendix B for more on this). Yet, these cases can, in our view, be dismissed on empirical grounds: we do observe (more than one) cities throughout our sample period. The exact model parameters involved depend, of course, on the specific model one considers.

[8] Scattered information on some of these variables is sometimes available, but it typically involves information for one particular city in one particular year. Comprehensive data, covering all countries and centuries in our sample, is, to the best of our knowledge, unavailable except for overall population.

[9] Table A1 in Appendix A provides descriptive statistics on all variables discussed in this section.

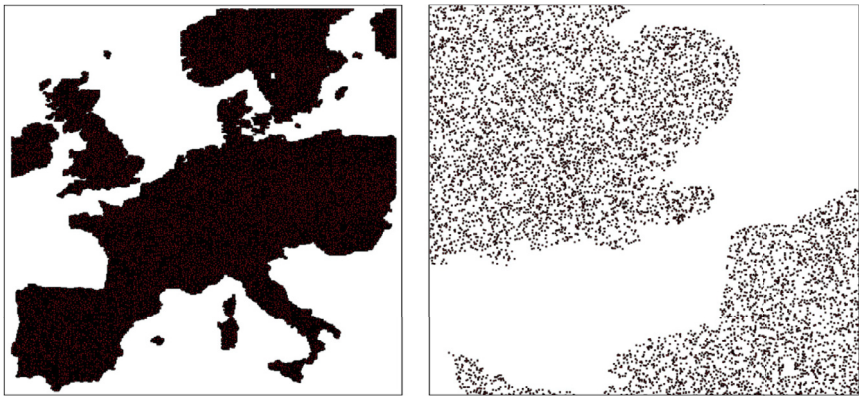
[10] We draw each location's latitude and longitude independently from a uniform distribution on the interval [36.02, 63.52] and [−9.38, 23.25] respectively; where these bounds are set at the minimum and maximum latitude (resp. longitude) of the actual cities in our sample. Coordinates are rounded up to 3 digits behind the comma, hereby implicitly setting a minimum distance between sampled locations of about 100 m (given the Earth's curvature this minimum distance depends on a location's exact latitude and longitude). For this reason one could also interpret our empirical strategy as randomly sampling grid cells from an exhaustive grid of 100 m × 100 m grid cells, where we take the midpoint of each grid cell as representative for the entire cell. We actually drew 400,000 random locations. Of them, 259,776 were located on land, and in one of the countries that we consider.

[11] Using an entire grid of this precision is computationally too intensive (implying 3.5 billion observations given the 10 centuries in our sample). Using a coarser exhaustive grid consisting of say 10 km^2 grid cells would still amount to about 3.5 million observations. More importantly, the use of a coarser grid would, in our view, unnecessarily complicate the choice of variables. It would require making more (arbitrary) choices as to how to aggregate up our variables of interest. For example, location on a river or road, or elevation are more ambiguously defined for a 10 km^2 grid cell compared to the 0.01 km^2 grid cells we now (implicitly) consider. Moreover, for all of our 2nd nature geography variables one would have to make a choice on how to measure distance from a gridcell to e.g. the nearest city (do we take the midpoint, boundary, or population weighted center of the gridcell?). Finally, the use of a coarser, but exhaustive, grid approach would in our view also raise the question of how one exactly places the grid over the geographical area one is studying. This choice may not be trivial, moving it say 1, 2, or 5 km to the north-, east-, south- and/or west could easily result in different empirical findings. All these issues become more and more salient, the larger the gridcells used.

[12] Note that some of these boundaries appear somewhat artificial in that they "cutoff" a part of Ireland, Portugal (a very small part west of Lisbon), Finland, Sweden or Norway. Not considering these countries in their entirety was a direct result of our sampling strategy that uses the eastern-/western-/northern- and southern most existing city (during our sample period) as the boundaries of our sample. Do note that the omission of these areas does not pose any problem for our inference, unless one suspects that locations in these areas have some unobserved features that explain both their exclusion from our sample as well as city emergence [endogenous sample selection]; this is very unlikely in our view.

[13] Indeed, excluding the Nordic countries or all locations above 54 degrees north latitude (25% of our sample, roughly everything north of York in the UK) does not affect any of our findings [see Table 1 column 7].

[14] Also in archaeology, it is common practice to define cities as population centers with more than 5000 inhabitants. See for example Fagan (1997, p. 27) or Bahn (1996, p. 57).



(a) full sample (b) zoomed in: the English Channel

Fig. 2. Potential city locations. *Notes:* each dot represents a potential city location.

(e.g., city rights in one part of Europe are not necessarily directly comparable to those in other parts). An absolute population cutoff to define a city avoids these issues of comparability; it makes the city definition less subjective. Moreover, we can easily compare the results using different, even possibly time-varying, population cutoffs (in Table A2 in Appendix A we do just that), or check whether measurement error crucially affects our findings (see Table 1, column 9).

With our city definition in hand, we use Bairoch et al. (1988) as our main source of the location of actual cities.[15] They provide centennial population estimates for all places in Europe that in some century over the 800–1800 period have more than 5000 inhabitants.[16] In total, this gives us 1588 actual city locations. Next, we match each of these cities to its closest randomly selected coordinate pair. We only replace this closest randomly selected coordinate pair by the actual city if it lies within at most 2.5 km to that city.[17] This results in 1150 matches (72.4% of our 1588 actual cities). Note that under our random sampling strategy this loss of actual cities in principle only matters for inference by making our estimates less precise, not inconsistent.[18] In robustness checks we also use a stricter 1.5 km maximum matching distance.

In sum, our sample consists of 259,776 randomly selected potential city locations, of which 1150 (0.4%) actually develop into a city at some point during our sample period.

3.3. Explanatory variables determining city location

3.3.1. 1st nature geography

To capture a location's opportunities for water- and land-based transportation, we construct dummy variables that indicate whether or not it has direct access (i.e. located within 500 m) to the sea, to a navigable waterway (river or lake), or to the former Roman road network. Under the assumption that the Romans mostly build their roads along least-cost routes, we take location on a Roman road as a proxy for a location's land-based accessibility.[19] For all our potential city locations, the information on their distance to the sea, lakes and navigable rivers[20] is obtained using GIS maps of Harvard's Center for Geographic Analysis. The information on the presence of a Roman road comes from Talbert (2000), as digitized by the DARMC project at Harvard.[21] Besides documenting whether or not a location has direct access to a Roman road, we also classify locations where two (or more) Roman roads crossed as hub locations.

Besides these transportation related 1st nature geography variables, we collected information on each potential city location's elevation [in meters] and on its ruggedness [calculated as the standard deviation of the elevation of the terrain within 10 km of a potential city location]. Both serve as a proxy of a location's accessibility, but can also be argued to be related to its agricultural possibilities. This data is taken from the (1×1 km) GLOBE database made available by the US National Geophysical Data Center.[22]

Finally, we collected information on each location's agricultural conditions from Ramankutty et al. (2002). That study combines information on climatic conditions (surface air temperature, precipitation and potential sunshine hours) and soil quality (total organic content [carbon density], availability of nutrients [pH] and water holding capacity) into one index that gives the probability that a certain location can be cultivated. This data is available in gridded form at a resolution of 0.5 degrees latitude–longitude (in case of our sample this corresponds to a grid of on average 56 km by 39 km). We match each potential city location to this data on the basis of its coordinates. Locations falling within the same grid cell have the same cultivation probability.

[15] Although this data has been critized (see e.g. Malanima, 1998), it is the main source for most recent empirical contributions considering Europe's long-run urban development (DeLong and Shleifer, 1993; Acemoglu et al., 2005; Nunn and Qian, 2011; Bosker et al., 2013).

[16] There are no population estimates for 1100. For this century we linearly interpolated the reported 1000 and 1200 population estimates. All our results are fully robust to excluding these interpolated 1100 numbers from the analysis. Results available upon request.

[17] The median [average] distance of a randomly drawn coordinate pair to its nearest already existing city is 1.79 km [2.11 km].

[18] We do use the information on all 1588 cities when constructing the explanatory variables capturing a location's 2nd nature geography characteristics (see Section 3.3.2 for more detail).

[19] We take Roman instead of actual roads for two reasons. First, Roman roads, constructed using similar methods and adhering to uniform quality standards can be found throughout the formerly Roman parts of Europe. Second, using Roman roads avoids some of the reverse causality issues that could arise when using actual roads. Of course, it remains a much less direct measure of 1st nature geography than the other variables discussed in this subsection. Moreover, a substantial part of our sample did not fall under Roman rule, so that our measure does not do full justice to locations outside the former Roman Empire. Excluding our dummy variables related to location on the former Roman road network from the analysis, or focussing only on locations within the boundaries of the former Roman Empie, does not affect any of our results.

[20] Our river variable only considers the larger European waterways. We lack information on smaller waterways. These will have been less important for a location's transportation possibilities, but were still an important source of freshwater, sewage, and sometimes also power.

[21] See http://darmc.harvard.edu.

[22] See http://www.ngdc.noaa.gov/mgg/topo/globe.html.

Table 1
Main results.

	$P(city\ t \mid no\ city\ t-1)$	1 BASELINE	2 NUTS3/ century FE	3 Location FE	4 Matched 1.5 km	5 + Institutions	6 + Ever city before?	7 Latitude <=54N	8 No 1800	9 Measurement Error
X_i	sea	0.007***	0.007***	–	0.004***	0.007***	0.006***	0.013***	0.005***	1
		[0.00]	[0.00]	–	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	<1>
	river	0.012***	0.012***	–	0.007***	0.012***	0.010***	0.012***	0.008***	1
		[0.00]	[0.00]	–	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	<1>
	hub	0.069***	0.070***	–	0.050***	0.062***	0.057***	0.069***	0.053***	1
		[0.00]	[0.00]	–	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	<1>
	road	0.004***	0.004***	–	0.002***	0.004***	0.004***	0.004***	0.003***	1
		[0.00]	[0.00]	–	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	<1>
	ln elevation	+0.00	0.00002	–	0.00001	+0.00	−0.00001	0.0001	0.00003	0
		[0.98]	[0.57]	–	[0.53]	[0.99]	[0.86]	[0.14]	[0.17]	<0>
	ruggedness	0.0001***	0.0001	–	0.00003	0.0001***	0.0001***	0.0001*	+0.00	0.99
		[0.00]	[0.17]	–	[0.11]	[0.00]	[0.00]	[0.05]	[0.93]	<1>
	P(cultivation)	0.0002**	0.0001	–	+0.00	0.0002**	0.0002**	0.0002**	0.0002**	0.91
		[0.02]	[0.28]	–	[0.76]	[0.04]	[0.03]	[0.04]	[0.02]	<0.99>
X_{it-1}	city >= 10k? $(t-1)$									
	0–20 km	−0.0005***	−0.0004**	−0.0001	−0.0003***	−0.0004***	−0.0004***	−0.0005***	−0.0005***	1
		[0.00]	[0.02]	[0.79]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	<1>
	20–50 km	0.0002**	0.0002*	0.0002**	0.0001	0.0002**	0.0002**	0.0002**	0.0001	0.51
		[0.02]	[0.07]	[0.04]	[0.15]	[0.02]	[0.01]	[0.02]	[0.39]	<0.83>
	50–100 km	0.0002***	0.0002***	0.0002***	0.0001***	0.0002***	0.0002***	0.0002***	0.0001**	1
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.03]	<1>
	$P(city\ t \mid no\ city\ t-1)$ unconditional	0.0005	0.0005	0.0005	0.0003	0.0005	0.0005	0.0006	0.0003	–
	NUTS2/century FE	Yes	NUTS3/century	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Nr observations	2,588,903	2,588,903	2,596,010	2,590,374	2,588,903	2,588,903	1,935,857	2,330,271	2,588,903
	p-value [20–100 km > 0?]	[0.00]	[0.01]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.09]	–
	# NUTS regions	247	1167	247	247	247	247	216	247	247
	% w/ useful variation	23.6	8.1	23.6	15.8	23.6	23.6	26.0	18.5	–
	% obs. in these NUTS	28.9	11.5	28.9	20.5	28.9	28.9	34.7	23.8	–

Notes: p-values, based on standard errors clustered at the NUTS3-century level between square brackets. +0.00 or −0.00 denotes a very small but positive, resp. negative, coefficient. Only in this table, we add X_i and X_{it-1} to the left of our variable names to clarify the correspondence between Eq. (2) and our results. In column 6 we control for a potential city location's religious, political and educational status in period $t-1$. In particular we include three dummy variables indicating whether a location already had an important religious [archbishopric], political [capital], or educational [university] function in period $t-1$. The estimated coefficient on these variables is 0.24, 0.30 and 0.55 respectively, all significant at the 1% level. In column 7 we also include a dummy variable indicating whether or not a location had already classified as a city in an earlier century (i.e. before $t-1$). The estimated coefficient on this variable is 0.43 and is significant at the 1% level. In column 9 we report the fraction of simulations that each respective variable is significant at the 5% or <10%> level.

* Significance at the 10%.

** Significance at the 5%.

*** Significance at the 1%.

The Ramankutty et al. (2002) data provides a time-invariant indication of a location's agricultural possibilities. It is not unlikely that a location's agricultural conditions (and most notably its climatic conditions) varied over the centuries. To our knowledge however, historical climate data is not available at a sufficiently disaggregated spatial scale to be useful for our purposes. To overcome this difficulty we capture the possibly time-varying agricultural conditions by including *region-century* fixed effects in all our baseline model specifications that are based on the 247 NUTS2 regions (as classified by Eurostat) in our sample. These NUTS2 regions cover on average an area of about 120 × 120 km. Besides controlling for time-varying regional agricultural conditions, these *region-century* fixed effects also capture any region-specific institutional, political, demographic or economic developments that may have left their mark on locations' urban chances. Notably, they control for the general increase in overall population that European countries experienced (each at a different rate) during our sample period (McEvedy and Jones, 1979).

Allowing for unobserved time-varying, geographically clustered heterogeneity is particularly important to get accurate estimates of the effect of 2nd nature geography. A location that is located in a region that is, for unobserved reasons, a good seedbed for city development, will have a high probability of becoming a city. But, so do other locations in that region. As a result, this location is also more likely to be surrounded by some already existing cities. When not adequately controlling for geographically clustered unobserved heterogeneity, one can thus easily, and wrongly so, ascribe an important role to 2nd nature geography. In our baseline results the included *NUTS2-century* fixed effects serve to alleviate these concerns. In robustness checks we also use a geographically even more fine-grained region-century specification based on the 1067 NUTS3 regions in our sample (as classified by Eurostat, covering on average only about 55 × 55 km). Moreover, we show results when controlling for all unobserved time-*invariant* location-specific characteristics (including location fixed effects). This addresses the caveat that we only have information on a subset of locations' important physical geography characteristics. By including location fixed effects, we take all time-invariant physical geography characteristics into account.

3.3.2. 2nd nature geography

The most commonly used measure of a location's 2nd nature geography is its market or urban potential (see e.g. Stewart, 1947; De Vries, 1984; Black and Henderson, 2003; or Bosker et al., 2013). This measure is a distance weighted sum of the population of all other already existing cities. In each century t, city i's urban potential (UP) is calculated as follows:

$$UP_{it} = \sum_{j=1, j\neq i}^{N} \frac{pop_{jt}}{D_{ijt}}. \quad (1)$$

We argue that such UP-type measures do not do justice to theory when looking at the establishment of new cities. The way UP is constructed implies that the impact of 2nd nature geography diminishes linearly with the size of, and distance to, other already existing cities. Moreover, it assumes that the impact of an already existing city on a location's own urban chances is either always negative or always positive (depending on the sign of the estimated coefficient on UP).

This is clearly too restrictive when looking at what theory predicts (see Section 2.2 or Appendix B). An existing urban center exerts an urban shadow at close range, prohibiting the formation of new cities in its immediate neighborhood. At the same time, potential locations that are too far removed from an already existing city also have little chance of becoming a city. It is the locations at medium distance from an already existing city that have the best urban chances. Theory thus predicts that an existing city exerts a *non-linear* effect on its surroundings. UP-type measures fail to adequately capture this.[23]

To do more justice to theory, we adopt the following dummy variable approach that does not a priori restrict the effect of existing cities to be positive or negative at all distances.[24] We first draw three concentric circles around each potential city location at ever further distance,[25] i.e. 20 km, 50 km and 100 km respectively. These distances roughly correspond to a 1 day, 2.5 days, and 5 days round-trip during most of our sample period.[26]

Next, we construct three dummy variables that indicate whether or not we find at least one already existing city of at least 10,000 inhabitants[27] within each respective 0–20 km, 20–50 km and 50–100 km distance band.[28] Fig. 3 illustrates the construction of these dummy variables in case of a hypothetical potential city location A. For this location, the dummy variables indicating the presence of an established urban center only take the value 1 in case of the 20–50 km and the 50–100 km distance band (there are no already existing urban centers within 20 km of A).

Three notes are in place here. First, we use the information on all 1588 actual cities in our sample in constructing these dummies. Second, we take explicit account of the fact that the world does not end at the boundaries of our sample. We consider any existing cities with more than 10,000 inhabitants outside of our sample area when constructing these dummies. Information on these cities also comes from Bairoch et al. (1988).[29] Third, note that the first distance band is somewhat stylized, following from our spatial characterization of a city by a unique central coordinate pair. Cities are not points, but cover a larger area depending on their size. A city of 10,000 inhabitants roughly covers 1 km^2 (using the generally accepted measure of 100 people per hectare), so that the city boundary lies at about 0.5 km from its centroid. For a city of 250,000 inhabitants this distance from centroid to boundary would still be "only" 2.5 km (only London, Paris and Naples pass this threshold in our dataset). All our results hold up to changing the first distance band to 0.5–20 km or even 2.5–20 km by only considering locations at at least 0.5 km or 2.5 km from an already existing city as potential city locations.

[23] Using such UP-type measures in our empirical specification shows a negative effect of already existing cities on their surroundings that is stronger for locations nearer the already existing city (see Bosker and Buringh, 2013). These results are available upon request.

[24] It does constrain the effect to be the same within each distance band. But, one can experiment with different/more distance bands (see Fig. A2). We also show results of abandoning our distance band-approach altogether, and fitting a sixth order polynomial in distance to the nearest city to the data instead (see Section 6.1).

[25] We calculate great circle distances between all locations in our data set on the basis of their coordinates.

[26] Roughly, because this depends on the mode of transportation, travel on horseback or donkey was generally faster than travel by foot, cart or water.

[27] We construct the dummy variables on the basis of existing cities larger than 10,000 inhabitants instead of 5000 inhabitants to limit possible reverse causality (simultaneity) issues from including a spatially lagged variable (similar to Hanson (2005)). In Table 2, we also show results when constructing these dummy variables on the basis of a larger and/or smaller population threshold for existing cities.

[28] An even more elaborate way to refine our 2nd nature geography variables would be to take account of e.g. actual road or river systems, or the ruggedness of the terrain, and come up with more detailed indicators of travel times between locations. Aside from the additional data requirements, note that such extensions, unless travel times between all locations in all centuries were available, also require making assumptions on the relative importance of each of the additionally considered characteristics in determining overall travel times. We leave such extensions for future work.

[29] Kaliningrad (Russia), Rovinj (Croatia), Ljubljana (Slovenia), Sombor and Subotica (Serbia), Timisoara, Satu Mare and Arad (Romania), Lvov and Brody (Ukraine) all had more than 10,000 inhabitants in at least one century of our sample period, and are located close to the boundaries of our sample. During our sample period no cities larger than 10,000 inhabitants exist in the parts of Norway, Sweden, Finland, Portugal or Ireland that lie outside the boundaries of our sample.

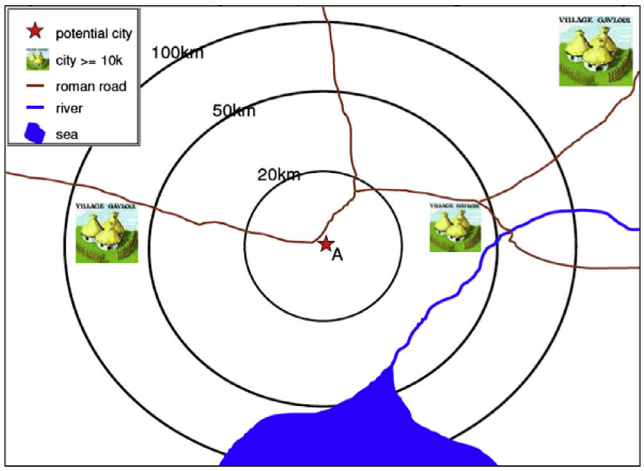


Fig. 3. Constructing dummy variables to capture 2nd nature geography.

4. Empirical framework

With our data in hand, we empirically quantify the effect of a location's 1st and 2nd nature geography characteristics on its urban chances, using the following simple empirical model:

$$P(c_{irt} = 1 | c_{irt-1} = 0, X_{irt-1}, X_i, \alpha_{irt}) = X_{it-1}\beta_1 + X_i\beta_2 + \varepsilon_{irt}, \qquad (2)$$

where c_{irt} is a dummy variable indicating whether or not location i in region r is a city in century t, X_i are our time-invariant 1st nature geography variables measured at the location level, and X_{it-1} are the three 'already existing city'-dummy variables capturing a location's 2nd nature geography that we introduced in Section 3.3.2. We consider these dummy variables lagged one century to further limit simultaneity issues resulting from the inclusion of these 'spatially lagged' variables (see also Footnote 27). ε_{irt} captures any unobserved/unmodelled variables at the city, region or century level affecting a location's urban chances. In our main specification we capture ε_{irt} by a full set of *NUTS2-century-specific* fixed effects and a location-century specific error term: $\varepsilon_{irt} = \alpha_{rt} + \upsilon_{it}$, where $r =$ NUTS2. In robustness checks we also show results using various different specifications for ε_{irt} (e.g. including location-specific century-invariant fixed effects, or taking $r =$ NUTS3). The β's are our parameters of interest. They reveal the sign, size and, together with their estimated standard error, significance of the included 1st and 2nd nature geography variables.

Our main empirical specification is essentially a (restricted) dynamic linear probability model[30]: we explain city presence conditional on city absence one century before. Given this dynamic specification, we need to assume no serial correlation in υ_{it} in order to obtain consistent estimates of our parameters of interest using standard techniques[31] (see e.g. Woolridge, 2005). We do allow for spatial correlation in υ_{it} within the same NUTS3 region, and base our inference on standard errors clustered at the NUTS3-century level.

5. Results

Table 1 shows our baseline results. When interpreting our findings, it is useful to always keep in mind that the average unconditional probability of becoming a city in one of the centuries in our sample is about 0.05%. This puts crucial perspective on the generally small magnitude of the estimated coefficients.

Column 1 shows our baseline findings: *1st nature geography is a very important determinant of a location's urban chances*. Especially preferential location for water- or land-based transportation substantially increases a location's probability to become a city (by 1.2 ppt (rivers) and 0.7 ppt (sea), or 0.4 ppt (road) and 7.3 ppt (hub) respectively).[32] Favorable agricultural conditions, or more rugged terrain also contributes positively to a location's urban chances. These effects are however much smaller, and much less robust (see e.g. columns 2, 4 or 8) than that of favorable location for river-, sea- or road-transportation. Corroborating Bairoch's claims, transport indeed played a crucial role in the location of cities.

Compared to 1st nature geography, *2nd nature geography's impact on a location's urban chances is smaller*. What is very interesting however, is that our flexible modelling strategy uncovers a pattern that is consistent with predictions made by economic geography theory (see e.g. Fig. B1 in Appendix B). The effect of an already existing city on another location's urban chances depends non-linearly on a location's distance to that city. Only locations at medium distance (20–100 km) from an already existing city have significantly better urban chances. They have about a 0.02 ppt higher probability to become a city than locations located further than 100 km away. Competition instead dominates at close range: an already existing city within 20 km significantly diminishes a location's own urban chances by about 0.05 ppt.[33]

The other columns of Table 1 show that the above-discussed baseline results hold up to a wide variety of robustness checks.[34]

5.1. Robustness: unobserved heterogeneity and matching locations to actual cities

Columns 2 and 3 verify whether our findings critically hinge upon capturing any possible unobserved time-varying geographically clustered heterogeneity by including NUTS2-century fixed effects in our baseline specification. Especially the estimated effect of 2nd nature geography could be sensitive to this, as one can

[30] In an earlier working paper version we relied on a probit specification instead, and also showed results using a logit, conditional random effects, or duration specification. Except for the estimated coefficient on our "hub" variable (that is much larger when using a linear probability specification – see also Footnote 32), our results do not depend on using either of these specifications (see e.g. Bosker and Buringh, 2013). The reason we now use a linear probability specification is that the non-linear nature of probit, logit or duration models does not easily accommodate the inclusion of all the 2470 NUTS2- or even 11670 NUTS3-century fixed effects. Apart from concerns about incidental parameter problems, it practically becomes computationally too intensive to calculate the average partial effect of each included regressor in (2) on a location's urban chances.

[31] Be it probit, linear probability, logit, or any other standard technique. Ways to relax this admittedly strong assumption have been proposed (Stewart, 2006 or 2007), but typically involve elaborate estimation techniques that rely on other, very specific, assumptions about the error term themselves. Note that the assumption of no serial correlation in the error term also ensures that we consistently estimate the effect of our one-century lagged 2nd nature geography variables (even in the presence of any remaining spatial correlation in the error terms).

[32] The effect of location on a hub of former Roman roads may appear to swamp the other effects. However, we do not want to stress the size of the hub effect for two important reasons. First the estimated hub effect is only this large when estimating a linear probability model. Using a logit, probit or duration specification, we also always find a significant hub effect, but one that is typically more than 10 times (!) as small as that when estimating a linear probability model. Taking the estimated hub effect of 7.3 ppt seriously, would result in severely overpredicting hub locations' urban chances in the data. The fact that the hub variable is the only variable whose estimated effect is this sensitive to using a linear probability model is not that surprising, given that hub locations are, by far, the least prevalent in our data set (see Table A1). On top of this, hub location is arguably the most endogenous of our first nature geography variables (see also Footnote 19). When building a new road that will cross an already existing road, an already existing city on the already existing road will be a likely focal point for this crossing (indeed, a Roman settlement existed on virtually all these crossings, see http://darmc.harvard.edu, so that part of our hub finding could be capturing a location's Roman urban history instead). Leaving out the hub variable altogether does not have any effect on any of the findings in this paper. They are available upon request.

[33] Note that all effects are relative to locations at more than 100 km from an already existing city.

[34] Table A2 in Appendix A further verifies the sensitivity of our baseline results to the absolute population cutoff of 5000 inhabitants that we use to define a city.

easily mistakenly take any remaining unmodelled geographically clustered heterogeneity for spillover effects.

In column 2, we use NUTS3-century fixed effects instead that are based on a division of Europe into 1167 NUTS3 regions instead of the larger 247 NUTS2 regions used in our baseline. Do note that using this more finegrained NUTS3-century specification results in a substantial loss of variation (only 8% of all NUTS3-century observations, containing 12% of all observations, provide useful identifying variation, versus 24% of all NUTS2-century observations, containing 29% of all observations[35]). Nevertheless, column 2 shows no substantial changes to our baseline findings: only the effect of a location's ruggedness and cultivation potential lose their significance. Given that these are the geographically most widely defined of our 1st nature geography variables (see Section 3.3.1) this is not that surprising.[36]

In column 3, we additionally control for any unobserved time-*invariant* factors at the *potential city location level* by including location-specific fixed effects. They do not only capture unmodelled 1st nature geography characteristics, but also control e.g. for a location's (urban) history *prior to our sample period*. A disadvantage of doing this is that we can no longer say much about the (relative) importance of specific 1st nature geography characteristics. However, it does allow us to verify the robustness of our 2nd nature geography findings to controlling for any of a location's (time-invariant) physical geography characteristics. Column 3 shows that our main 2nd nature geography results hold up to this important robustness check. Although the negative competition effect at short range (0–20 km) loses its significance, we still find the significant positive effect of being located at medium distance to an already existing city.

5.2. Robustness: additional variables and sample composition

In columns 5–6 we show the robustness of our findings to adding additional variables to our baseline specification. Reassuringly, and with only few exceptions, all our main baseline results come through.

Column 5 controls for a potential city location's religious, political and educational status in period $t-1$. These data are taken from Bosker et al. (2013). We find that having an important religious [archbishopric], political [capital], or educational [university] function substantially increases a location's urban chances (see the *Notes* to Table 1). The results on these variables should be taken with some care. Only 28 (or 0.001%) of all our potential city locations are a capital or have a university, and only 75 (or 0.003%) are an archbishopric, before becoming a city. Although these characteristics significantly improve a location's urban chances, such locations are major exceptions.

Column 6 instead controls for a location's (urban) population history. We include a dummy variable indicating whether or not a location had ever been a city before. This is done to control for the presence of cities (0.02% of the sample) that at some point pass our 5000 inhabitants criterion, subsequently fall back below this number, to pass it again in a later century.[37] These cities would – so to speak – be counted double in our sample, which could leave an effect on the results. This is however not the case. Locations that once already qualified as a city do have a 43 ppt (!) higher probability to regain their city status.[38]

Next, columns 7 and 8 show that our results are not driven by (implicit) choices regarding the spatial and temporal reach of our sample. Column 7 shows that our findings are not driven by the implicit oversampling of locations are higher altitudes (see our discussion in Section 3.1). Excluding all locations above 54N latitude (25% of our sample) does not affect any of our results.[39] Column 8 instead shows the insensitivity of our main results to excluding 1800 from the sample. As shown in Fig. A1 in Appendix A, the eighteenth century saw an unprecedented increase in the number of cities. Column 4 shows that it is not only this episode that drives our results.[40]

5.3. Robustness: measurement error

The final robustness check deals with the important issue of measurement error. Bairoch et al. (1988) acknowledge that getting spot-on population estimates is sometimes difficult, especially for the smaller cities and for the earlier centuries. As we are using a nonlinear transformation of the city population data, such measurement error, even if it were random, could leave its effect on our results (see e.g. Hausman, 2001). To verify the sensitivity of our results to measurement error, we adopt the following simulation strategy.

We assume that each reported population estimate has a 40% probability of being misreported. Conditional upon being misreported, we subsequently assume that there is an equal, 25%, chance of being underestimated by 2000 inhabitants, overestimated by 2000 inhabitants, underestimated by 1000 inhabitants, or overestimated by 1000 inhabitants respectively. This structure for the measurement error implicitly assumes that Bairoch et al. (1988) made relatively bigger mistakes for smaller population numbers.[41] We generate 1000 different population samples using this sampling strategy and estimate our baseline model using each of these samples separately. Column 9 reports the fraction of simulations that each variable is significant at a 5% and <10%> level respectively. Under the assumption of measurement error, each of the 1000 simulated samples is 'equally true'. If we find that a significant variable in our main results is significant at the 10% level in less than 90% of our simulations, this would shed some doubts on the actual relevance of this variable. Reassuringly, all our results hold up to this measurement error check. The significant positive effect of an already existing city at a distance of 20–50 km is most sensitive to this measurement error check. However, it is still significant at the 10% level in 83% of our simulation runs.

Summing up all results discussed in this section, they show that 1st nature geography is the dominating factor in determining locations' urban chances. Preferential location for water- or land-based transportation is a particularly important city seed. Second nature geography also matters, but to lesser extent. Interestingly, we do find evidence that an already existing city influences the urban chances of its surroundings in a way that is consistent with predictions from all recent new economic geography models. Only an already existing city at medium range raises a location's urban chances.

[35] A region provides useful identifying variation in a particular century only if at least one city appears in that region in that century (otherwise the region-century dummy perfectly predicts the non-appearance of a city for each and every location in that region). When looking over the entire sample period, about 85% of all NUTS2, but only 50% of all NUTS3, regions ever provide useful variation.

[36] All main results in our paper are qualitatively robust to using a NUTS3-century specification. Given that we lose a substantial amount of variation when using such a NUTS3-century specification, it does make some of the estimates in our extensions less precise (point estimates do always indicate the same patterns).

[37] The Black Death (the 14th-century plague epidemic) is responsible for many of these 'city-disappearances'. 40% of the existing cities in 1300 'disappeared' in 1400, i.e. fell back below the 5000 inhabitants threshold.

[38] This effect is substantially smaller when using a probit, logit or duration specification.

[39] Similarly, excluding all Nordic countries from the sample also does not change any of our findings.

[40] Our results are also robust to excluding the UK, the earliest industrializing country, from the sample.

[41] Bairoch (1988, p. 525) expects a margin of error of about 10% for overall European city population around 1300 and 1500, increasing to 15% in 1000 and even 20% in 800.

However, these 2nd nature findings are most vulnerable to any remaining omitted variable concerns. We cannot fully exclude the possibility that they are in fact driven by an unobserved, time-varying, spatially clustered, 1st nature geography characteristic at a spatially even more disaggregated level than that of the 1167, on average 55 × 55 km, NUTS3 areas in our sample.[42] In the next section we therefor show several extension of our main findings that aim to provide additional confidence in our 2nd nature geography results in particular.

6. Second nature in more detail

We refine our 2nd nature geography results in two ways. First, we show results of using several *spatially more elaborated* 2nd nature geography specifications. Second, we show how the effect of both 1st and 2nd nature geography on a location's urban chances *changes over the centuries* in our sample. Economic geography theory makes predictions under what conditions we should expect to find stronger 2nd nature geography effects, and at what distances to already existing cities (see the Theory Appendix). The patterns we find in our extensions can be readily compared to these predictions from theory. If we find patterns that are hard to reconcile theoretically this would certainly shed some doubt on our main results.

In addition to the results shown in this section, Section A.3 in Appendix A addresses the important concern that our 2nd nature geography results may simply be an artefact of the increased density of the European city system over the centuries. In later centuries there are fewer potential city locations left. But, those that are left are also more likely to be located closer to an already existing city than in earlier centuries. This could have unwanted consequences for our inference. Results from a Dartboard Approach in the spirit of Ellison and Glaeser (1997) and Duranton and Overman (2005) alleviate this concern.

6.1. More/different spatial detail

In column 1 of Table 2 we start by allowing the impact of an already-existing city to differ depending on its size. To do this we include six additional dummy variables to our baseline specification: three indicating the presence of *at least one already existing city larger than 25,000 inhabitants*, and three indicating the presence of *at least one already existing city larger than 5000 inhabitants* in each of the three distance bands respectively.

The results show an interesting pattern. We find that the larger the distance between an existing city and a potential city location, the larger the existing city has to be to exert a positive influence on that potential city location's urban chances. Put differently, the larger an already existing city, the larger its urban shadow (a finding that corresponds nicely with early observations by e.g. Lösch (1940, p. 126) or Ullman (1941, p. 856), and with predictions from new economic geography theory[43]). A city of 5000–10,000 inhabitants is simply too small to significantly affect the urban chances of other locations (the point estimates do show a similar non-linear pattern as for the effect of larger already existing cities). By contrast, the effect of a city larger than 10,000 inhabitants is positive at medium distances (20–100 km), and it carries significant negative effects at close range (0–20 km) – see also the *p*-values at the bottom of Table 2.[44] An even larger city of more than 25,000 inhabitants carries effects that are similar to those of a city larger than 10,000 inhabitants, except for the fact that its urban shadow is stronger: it only significantly positively affects the urban chances of locations beyond 50 km of the city itself.

Column 2 shows a very similar pattern. There we include three additional dummy variables to our baseline specification that indicate the presence of *at least two cities larger than 10,000 inhabitants* in each of the three distance bands respectively. As in our baseline findings, the presence of only one city larger than 10,000 inhabitants exerts a positive influence on the urban chances of locations at 20–100 km, and poses a negative urban shadow on locations at closer distance. A second already existing city does not further increase this positive effect, except when it is located sufficiently far away (at least 50 km).[45]

As a next spatial extension, we completely abandon our 'distance-bands approach', and model the possibly non-linear effect of 2nd nature geography by including a sixth order polynomial in *distance to the nearest already existing city* to (2) instead:

$$P(c_{ict} = 1 | c_{ict-1} = 0, X_{ict-1}, X_i, \alpha_{ict}) = \sum_{m=1}^{6} \left(\left(D_{it-1}^{near} \right)^m \gamma_m \right) + X_i^{GEO1st} \beta_1 + \alpha_{ict} + \varepsilon_{it}, \quad (3)$$

where D_{it-1}^{near} denotes the distance to the nearest already existing city from location i in century $t-1$, and X_i^{GEO1st} the same 1st nature geography variables as in our baseline model (see Table 1). Based on the six estimated γ_m's we can verify whether, and if so how, the marginal effect of distance to the nearest already existing city varies at different distances.

Fig. 4 shows the results (also showing 5% and 10% confidence intervals).[46] They again clearly confirm our baseline findings. Interestingly, we find a picture that bears a striking resemblance to the theoretical predictions illustrated by Fig. B1 in Appendix B. Being located further away from an already existing city significantly increases a location's urban chances up until 75 [66] km at the 10% [5%] level. Again evidence of an "urban shadow" effect. This turns around at 107 km from the nearest already existing city, and at distances between 136 and 399 km [144–377 km], we find a significant negative effect of being located further away from an already existing city.[47] Finally, at further distances this significant

[42] Do note that thinking of a geographically clustered omitted variable that affects city emergence differentially over the centuries is one thing. In order to completely explain our 2nd nature geography findings however, this variable would have to affect locations' urban chances differently (sometimes negatively and sometimes positively) depending on their distance to already existing cities. Thinking of such a variable is, in our view, much harder than of a variable that affects city emergence either always positively or always negatively. One possibility could be an omitted natural resource that became more important over the centuries ànd whose deposits always occur at 20–100 km from the next deposit. This would however still not explain the *significant* negative "urban shadow" effect at close distance to an already existing city. We have not found evidence of any natural resource whose deposits are scattered according to such a fixed spatial pattern.

[43] See e.g. the discussion around Fig. A2 in Fujita et al. (1999).

[44] Note the way the different dummy variables are specified: e.g. if there exists a city larger than 25,000 inhabitants within a certain distance band, not only the dummy variable indicating the presence of a city larger than 25,000 inhabitants will be 1, so will be the dummy variable indicating the presence of a city of at least 10,000 inhabitants, as well as the dummy variable indicating the presence of a city of at least 5000 inhabitants. The *p*-values below the coefficients therefor indicate whether or not the effect of an already existing city is significantly different from that of a smaller already existing city within the same distance band. At the bottom of Table 2, we therefor also show the *p*-values indicating whether or not the effect of having a city of 10,000 or 25,000 inhabitants is significantly different from zero for each of our three distance bands.

[45] Do note that it hardly ever happens that a location lies within 0–20 km of two already existing cities.

[46] The results on our 1st nature geography variables in the regression underlying Fig. 4 are very similar to those shown in Table 1 – column 1. They are available upon request.

[47] The distances at which the different effects are significant are generally larger compared to those found using the distance band approach. A crucial difference with our baseline distance band approach is that the 'polynomial-appraoch' only explicitly considers each location's nearest already existing city (i.e. *only one per location*), whereas the distance band approach is able to also take characteristics of the already existing urban system into account that go beyond each location's nearest already existing city. E.g. for a potential city location that is located at 15 km to

Table 2
2nd nature geography – some extensions.

P(city t|no city $t-1$)	1	2
1st nature geography as in Table 1		
	city >= 5k, $t-1$?	1 city >= 10k, $t-1$?
0–20 km	−0.0003	−0.0005***
	[0.17]	[0.00]
20–50 km	0.0001	0.0002**
	[0.340]	[0.03]
50–100 km	−0.00002	0.0002***
	[0.72]	[0.00]
	city >= 10k, $t-1$?	>=2 cities >= 10k, $t-1$?
0–20 km	−0.0001	0.0005
	[0.75]	[0.59]
20–50 km	0.0001	0.0001
	[0.45]	[0.39]
50–100 km	0.0002**	0.0002**
	[0.03]	[0.02]
	city >= 25k, $t-1$?	
0–20 km	−0.0005*	–
	[0.08]	–
20–50 km	0.00003	–
	[0.86]	–
50–100 km	0.0001	–
	[0.20]	–
NUTS2/century FE	Yes	Yes
Nr observations	2,588,903	2,588,903
p-values tests	H0: $\beta_{\text{city >= 10k}} > 0$?	H0: $\beta_{\text{2 cities >= 10k}} > 0$?
0–20 km	[0.04]**	[0.99]
20–50 km	[0.03]**	[0.06]*
50–100 km	[0.00]***	[0.00]***
	H0: $\beta_{\text{city >= 25k}} > 0$?	
0–20 km	[0.00]***	–
20–50 km	[0.105]	–
50–100 km	[0.00]***	–

Notes: p-values, based on standard errors clustered at the NUTS3/century level between square brackets. All regressions contain the same 1st nature geography variables as in column 1 of Table 1. The estimated parameters on these variables correspond closely to those reported there. They are available upon request.

* Significance at the 10%.
** Significance at the 5%.
*** Significance at the 1%.

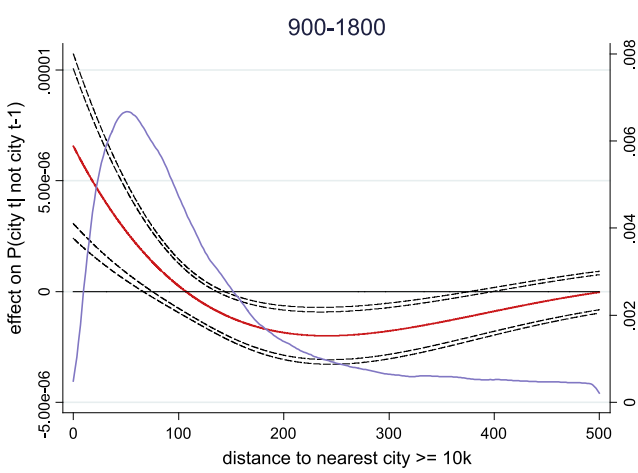


Fig. 4. Estimated 6th order polynomial in distance to the nearest city. *Notes:* distances are in kilometers. The gray line depicts the distribution of distance to the nearest city over all locations in our sample. The 5%-, 25%-, 50%-, 75%-, and 95% quantile of this distributions is 23 km, 62 km, 122 km, 357 km and 1011 km respectively. The dotted lines show 5% and 10% confidence intervals around the estimated marginal effect at each distance to the nearest city. They are calculated using standard errors clustered at the NUTS3/century level.

effect disappears again; but do note that only very few locations find their nearest already existing city at those distances [see the plotted distribution in, and *Notes* to, Fig. 4].

6.2. More detail over the centuries

Next, we extend our baseline results by looking for possible changes in the importance of 1st and 2nd nature geography over time. Any such changes are of course interesting in and of themselves. But, the results in this section, as those in Section 6.1, are aimed at instilling further confidence in our 2nd nature geography findings in particular.

As set out briefly in Section 2.2, and illustrated in more detail in Appendix B, economic geography theory predicts a positive effect at medium distance from an already existing city that depends delicately on the relative strength of the agglomeration and dispersion forces in the model, e.g. the size of trade costs, the extent of scale economies in urban production, and overall population size. Unfortunately, comprehensive data on the exact way in which these variables changed over our sample period is unavailable. Nevertheless, there is little discussion about the general trend in these variables over the centuries in our data. Improved transportation technology, but also the introduction of e.g. the bill of laden, and insurance contracts (Greif, 2006, p. 24) almost certainly reduced trade costs.[48] Second, if anything, scale economies in urban production increased due to improved non-agricultural production techniques (e.g. the blast furnace, finery forge, and the printing press). Moreover, the diseconomies of scale associated with urban life decreased over our sample period as a result of better sanitation, disease control, etc. All these developments would point to 2nd nature geography becoming more important over the centuries. If anything, we would therefor expect the positive effect at medium distance to show up most clearly in the later centuries in our sample.

Table 3 shows the results of estimating such a finer century specification, where we allow the effect of each variable to differ by 200 year periods.[49] Some care is warranted in interpreting the results for 1100–1200. They are based on the interpolated population numbers for 1100 (see also Footnote 16), both in determining city emergence in 1100 and 1200, as well as in constructing the three lagged already existing city dummies in 1200.

There is little evidence that the dominant role of 1st nature geography changed markedly over our sample period. Preferential location on the main transportation arteries in particular, is the dominant driver of city location throughout our sample period. Its dominance over the other 1st nature geography variables related to elevation, ruggedness or cultivation potential only increases over the centuries. A finding that is consistent with the notion that in later centuries food could be transported over greater distances at lower costs due to improvements in transportation, hereby diminishing the importance of location right next to fields of high agricultural productivity (Duranton, 1999, p. 2173).[50] Also, there is

an already existing city, at 40 km from another existing city, and at 95 km to a third already existing city, all three of these already existing cities would be taken into account when using our distance band approach, whereas only the closest would be taken into account in the 'polynomial approach' underlying Fig. 4.

[48] It is less clear how these improvements affected the relative size of trade costs for agricultural and manufacturing goods respectively (which is what matters according to theory).

[49] Of course it is also possible to estimate each coefficient by century. Doing this shows a very similar, yet much noisier pattern. Especially in the earlier centuries in the sample, such a century-by-century specification would leave us with only little useful identifying variation (exemplified by the much smaller unconditional probability of becoming a city in these earlier years).

[50] In north-western Europe e.g. the grain trade with eastern Europe became increasingly important (Hybel, 2002).

Table 3
A finer century decomposition.

P(*city t*|*no city t* − 1)	900–1000	*1100–1200*	1300–1400	1500–1600	1700–1800
sea	0.002*	*0.001*	0.005***	0.010***	0.019***
	[0.07]	*[0.16]*	[0.00]	[0.00]	[0.00]
river	0.002***	*0.003****	0.011***	0.014***	0.034***
	[0.00]	*[0.00]*	[0.00]	[0.00]	[0.00]
hub	0.027***	*0.031****	0.055***	0.094***	0.176***
	[0.00]	*[0.00]*	[0.00]	[0.00]	[0.00]
road	0.001***	*0.001****	0.003***	0.005***	0.012***
	[0.01]	*[0.00]*	[0.00]	[0.00]	[0.00]
ln elevation	0.00001	*−0.00001*	−0.00001	0.0002***	−0.0002
	[0.6]	*[0.8]*	[0.8]	[0.00]	[0.13]
Ruggedness	−0.00004	*−0.00002*	0.0001	−0.0001	0.0005***
	[0.12]	*[0.48]*	[0.16]	[0.29]	[0.00]
P(cultivation)	0.0001	*+0.00*	0.0002	0.0004**	0.0004
	[0.32]	*[0.77]*	[0.24]	[0.04]	[0.21]
city >= 10k? (*t* − 1)					
0–20 km	−0.0003***	*−0.0002***	−0.0006***	−0.0008***	−0.0007**
	[0.00]	*[0.05]*	[0.00]	[0.00]	[0.05]
20–50 km	+0.00	*0.0001*	−0.00002	0.0002	0.0002
	[0.99]	*[0.31]*	[0.80]	[0.18]	[0.19]
50–100 km	+0.00	*0.0001**	+0.00	0.0002**	0.0005***
	[0.37]	*[0.06]*	[0.93]	[0.04]	[0.00]
P(*city t*|*no city t* − 1) unconditional	0.00006	*0.00008*	0.0003	0.0006	0.0015
NUTS2/century FE			Yes		
Nr observations			2,588,903		
p-value [20–100 km > 0?]	[0.66]	*[0.106]*	[0.89]	[0.06]	[0.01]
# NUTS regions	247	*247*	247	247	247
% w/ useful variation	2.6	10.3	22.4	30.6	52.2
% obs. in these NUTS	4.1	14.2	26.2	40.7	59.1

Notes: *p*-values, based on standard errors clustered at the NUTS3/century level between square brackets. +0.00 or −0.00 denotes a very small but positive, respectively negative, coefficient. Results for 1100–1200 are italic as these should be taken with some care (see the discussion in the main text).
* Significance at the 10%.
** Significance at the 5%.
*** Significance at the 1%.

some (be it much weaker) evidence of an increased relative importance of water- over land-based transportation.[51]

Turning to the changes in the importance of 2nd nature geography: they do instill further confidence in our baseline findings. Throughout the centuries, we find a strong and significant urban shadow effect. Locations within a 20 km range of an already existing city have significantly smaller urban chances. By contrast, the significant positive effect of an already existing city at medium distance appears to be driven by the later centuries. In the early centuries, we find no evidence that an already existing city exerts a positive influence on other locations' urban chances at any distance. Only from around the 16th century onwards we start to find that locations at medium distance from already existing cities have significantly higher urban chances than those at closer or further distance. This effect shows most strongly for locations at 50–100 km from an already existing city, but do note that a test for the joint significance of the effect of an already-existing city on locations at 20–50 km and 50–100 km is also always significant in the later centuries.

The emergence of this positive effect of 2nd nature geography in the later centuries only is very much consistent with theory (Behrens, 2007; Fujita and Mori, 1997; Duranton, 1999).[52] Ideally, we would be able to ascribe its emergence to changes in the important factors stressed by theory that affect the balance of agglomeration and dispersion forces. As mentioned before, comprehensive data on most of these variables is however unavailable. As a result, we can only say that the patterns in the data are consistent with theory. We cannot unambiguously ascribe them to (changes in) one or more of the factors stressed by theory as possibly explaining this pattern. One exception here is population. Data on total population of the countries in our sample is available over our sample period (McEvedy and Jones, 1979).

According to theory, the likelihood of finding a positive effect of location at medium distance from existing cities should increase with overall population size. Moreover, the range of distances at which the existing city exerts its positive influence should increase as population grows further, also expanding toward the already existing city (see Fig. B1). Table 4 shows that, if one is willing to abstract from changes in the other important agglomeration and dispersion forces,[53] our findings could in principle be explained by Europe's overall population dynamics. To do this we interact each of our three distance band dummies with total (urban + rural) population in the countries in our sample.[54] Note that the effect of overall population itself is perfectly captured by the included NUTS2/century fixed effects.

As predicted by theory, an increasing overall population significantly increases the urban chances of all locations, irrespective of

[51] This corresponds nicely to Lopez (1956) or Pirenne (1925). It also concurs with the notion that improvements in shipping technology (not only in the size and speed of the vessels used, but also in e.g. navigation (van Zanden and van Tielhof, 2009) and canal building (Bairoch, 1988)) were larger than those in land transportation despite the fact that e.g. horseshoes, rigid tandem horse collars, and the use of explosives to build tunnels, did all improve land-based transportation (Lopez, 1956). We do not want to emphasize this too much however, given that our land transport variables focus on the Roman road network only. See also Footnotes 19 and 32.

[52] The exact same pattern also shows up when modelling 2nd nature geography in a different way than our "three-distance-band-approach". See Fig. A3 and Table A4 in Appendix A.

[53] In reality it is not unlikely that the development of these variables over the centuries is correlated with that of overall population.

[54] Results are similar when using population distinguished by current-day European country instead. They are available upon request. Given that mobility in Europe is very low across national boundaries, one could argue that national population is what matters instead of overall European population. However, most of the national boundaries used by McEvedy and Jones (1979) in collecting their national population estimates, were never in effect throughout our sample period.

Table 4
Overall population and the impact of 2nd nature geography.

$P(city\ t \mid no\ city\ t-1)$	
1st nature geography: see Table 1	
city >= 10k? $(t-1)$	
0–20 km	−0.0014***
	[0.00]
20–50 km	−0.0005***
	[0.003]
50–100 km	−0.0004***
	[0.002]
city >= 10k? $(t-1)$ interacted with overall European population (in mln)	
0–20 km	0.00002**
	[0.021]
20–50 km	0.00001***
	[0.002]
50–100 km	0.00001***
	[0.00]
NUTS2/century FE	Yes
Nr observations	2,588,903

Notes: *p*-values, based on standard errors clustered at the NUTS3/century level between square brackets. All regressions include the same 1st nature geography variables as in Table 1. Their estimated coefficients are very similar to those reported in column 1 of Table 1, and are available upon request.
* Significance at the 10%.
** Significance at the 5%.
*** Significance at the 1%.

their distance to the existing city. Based on our estimates, we can also calculate how large overall population needs to be, to turn the negative effect of the existing city into a positive one. At the 5% [10%] level this happens first at a distance of 50–100 km of the existing city at an overall European population of 55.75 [52.5] million (from 1500 onwards). At 20–50 km from the existing city this happens later at a total population of 71.75 [67.25] million (from 1600 onwards). At closest range to the existing city the effect is always negative (and significantly so at the 5% [10%] level up to an overall European population of 47.5 [54.25] million). This pattern is also consistent with theory: when population is too small, no new city can emerge at any distance from the already existing city. As population increases, locations at medium distance from the already existing city become viable new city locations (here this happens first within the 50–100 km range). And, as population grows further, this positive range expands, also toward locations nearer the already existing city (here to places located within 20–50 km of the existing city).

7. Conclusions

This paper empirically disentangles the different roles of geography in determining the location of European cities. We introduce a new data set covering over 250,000 randomly drawn potential city locations, as well as all actual European cities. This data is available over the 800–1800 period, when the foundations for today's European city system were laid. Using this data, we uncover the importance of physical, 1st nature, geography and of the urban system already in place (2nd nature geography) in determining locations' chances of becoming a city. The empirical strategy we employ to do so, aims to alleviate the inherent difficulty of separately identifying 2nd nature geography from any remaining unobserved spatially clustered 1st nature geography as much as possible. Nevertheless it is important to keep in mind that our 2nd nature geography findings in particular hinge upon the validity of the specific identification assumptions that we make in our estimations.

Our results, that hold up to a wide-variety of robustness checks and extensions, show that geography indeed played a crucial role in laying the foundations of today's European cities. First nature geography is the dominant determinant of city location throughout our sample period. Preferential location for water- or land-based transportation is a particularly important city seed. And, its importance only increased over our sample period relative to other 1st nature geography characteristics, notably a location's own agricultural potential. A finding that is consistent with the notion that in later centuries food could be transported over greater distances at lower costs due to improvements in transportation, hereby diminishing the importance of location right next to fields of high agricultural productivity.

Second nature geography is less important, but, we show that the way it matters corresponds closely to predictions from (new) economic geography theory. We always find strong evidence of an urban shadow effect: throughout our sample period, already existing cities prohibit the development of other cities in their immediate backyards (within a range of about 20 km or a one-day round trip). This is very different at medium distance from an already existing city: locations at 20–100 km from an already existing city have significantly higher urban chances. These locations combine the advantage of cheaper trade with the existing city compared to locations at further distance, with that of weaker competition with the existing city compared to locations at closer distance. This positive effect at medium distance is particularly prominent in the later centuries in our sample. And, we show that Europe's overall population dynamics can in principle explain the emergence of this positive effect over our sample period.

Acknowledgments

We thank three anonymous referees and Edward Glaeser for comments and suggestions that have significantly improved this paper. Also we thank, Rob Alessie, Bob Allen, Bas van Bavel, Francesco Billari, Christiaan van Bochove, Marius Brülhart, Gilles Duranton, Jessica Dijkman, Rui Esteves, Giovanni Facchini, Oliver Falck, Ewout Frankema, Harry Garretsen, Knick Harley, Laura Hering, Tommy Murphy, Nathan Nunn, Rick van der Ploeg, Maarten Prak, Joppe de Ree, Roger Smeets, Tony Venables, Wouter Vermeulen, Joachim Voth, Bastian Westbrock, Nikolaus Wolf, Jan Luiten van Zanden, and participants at various workshops and conferences for their feedback on our paper. Special thanks go to Giovanni Zambotti at Harvard's Center for Geographic Analysis for helping us a lot in obtaining the necessary information for all our randomly drawn coordinate pairs.

Appendix A. Additional results

A.1. Descriptives

See Figs. A1 and A2 and Table A1.

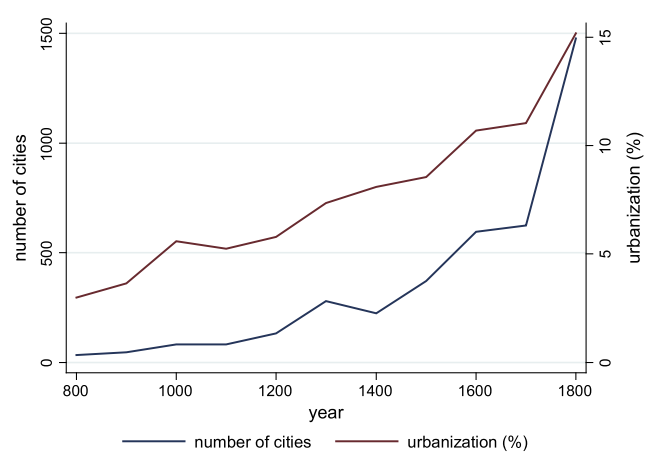


Fig. A1. Urbanization and the number of cities in Europe, 800–1800. *Notes:* Both the number of cities and the urbanization rate are based on defining cities as population centers with at least 5000 inhabitants [see Section 3.2 for more detail on this definition]. The urbanization rate is calculated by dividing total urban population (i.e. the total number of people living in cities with at least 5000 inhabitants) by total population. Total population figures are taken from McEvedy and Jones (1979).

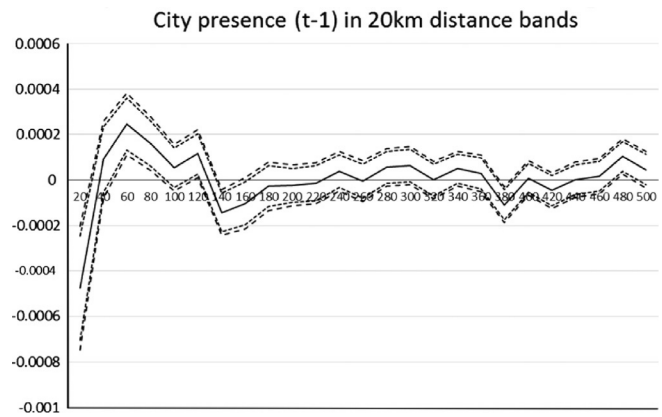


Fig. A2. Twenty-five 20 km distance bands (or 1-day round trip bands). *Notes:* This figure shows the results of estimating (2) including twenty-five already-existing city dummies instead of the three that we use in our main specification. Each of these twenty-five dummies indicates whether or not an existing city was present in century $t-1$ in respective 20 km distance bands up to 500 km from a potential city location. 20 km roughly corresponds to a one day round-trip during most of our sample period. The numbers on the x-axis denote the upper boundary of each distance band, i.e. 20 denotes the 0–20 km band. The dotted lines corresponds to the 5% and 10% confidence intervals around the estimated coefficient. They are calculated based on standard errors clustered at the NUTS3/century level. The underlying regression contains the same 1st nature geography variables as in column 1 of Table 1. The estimated parameters on these variables correspond closely to those reported there. They are available upon request.

A.2. Changing the city definition

It this section we assess the sensitivity of our main results in Table 1 with respect to our city definition based on an *absolute* population cutoff of having at least 5000 inhabitants. Table A2 shows the results when using a different absolute cutoff, or a time-varying population cutoff instead.

In columns 1 and 2, we lower our absolute population criterion to 3000 and 4000 inhabitants respectively. Bairoch et al. (1988) only provide population numbers smaller than 5000 inhabitants for a very limited set of city locations. They stress (see their p. 218) that these numbers are on the one hand subject to a much greater margin of error than those larger or equal than 5000 inhabitants, and, on the other hand, that they did not systematically search for any numbers smaller than 5000, so that these numbers are only very selectively available. Nevertheless, using these much less reliable population numbers, we find the same results as when using our preferred 5000 inhabitants cutoff to define a city.

When instead raising the population cutoff, all but one of our main results come through. When increasing our population cutoff to 6000 inhabitants, we find a slight change to our 2nd nature geography results that is further exacerbated when increasing the population cutoff to 10,000 inhabitants. In particular, we find that the positive effect of having an already existing city at 20–50 km disappears when raising the criterion to more than 6000 inhabitants. Do note however that a test for the joint significance of the 20–50 km and 50–100 km is always significant. Also, raising the cutoff further reduces the percentage of our observations providing us with useful variation to identify our effects of interest as fewer locations are now turning into a city.

However, this result does not necessarily invalidate our baseline results. In combination with our baseline findings, the results in columns 2–6 show a consistent pattern: the positive effect of an already existing city at medium distances gradually disappears when raising the absolute size criterion used to define a city. Having an existing city at 20–50 km may significantly improve a location's probability of becoming a city of 5000 or 6000 inhabitants, it becomes increasingly difficult to grow larger, say 10,000 inhabitants, in the shadow of an already existing urban center. An existing city, as it were, does only tolerate moderately sized new cities to appear in its immediate backyard. These results hereby nicely complement our findings in column 1 of Table 2 where we used different population thresholds when constructing "already existing city within x km"-dummy variables.

Finally, column 7 shows results when using a time-varying population cutoff to define a city. We employ the following step-wise increasing population cutoff: 5000 inhabitants before 1600, 6000 in 1600 and 1700, and 10000 in 1800. We choose this particular stepwise increase as it leaves the unconditional probability of becoming a city in any century around 0.03% (instead of increasing substantially over this period when using our absolute 5000 inhabitants cutoff). Using such a time-varying definition is in itself not without difficulties. In particular, given that we condition on not already being a city in $t-1$, one has to choose which definition to use when constructing these variables (i.e. the 'new' definition in period t or the 'old' definition in period $t-1$). In column 7 we use the city definition in period $t-1$ as our conditioning variable (i.e. was there a city in period $t-1$). The results show that our main findings are also robust to using this time-varying city-definition.

Table A1
Descriptives.

1st or 2nd nature characteristic	Mean	sd	min	max	Mean	sd	min	max
	All locations (259,776)				Locations ever >= 5000 (1150)			
sea	0.006	0.08	0	1	0.09	0.28	0	1
river	0.02	0.13	0	1	0.48	0.50	0	1
hub	0.001	0.03	0	1	0.13	0.34	0	1
road	0.02	0.16	0	1	0.36	0.48	0	1
elevation (m)	382	429	−15	4356	232	242	−3	1218
ruggedness	90.2	110.6	0	922.8	79.0	86.6	0	720.8
P(cultivation)	0.53	0.34	0.001	0.999	0.71	0.23	0.006	0.999
Latitude	50.0	6.8	36.0	63.5	46.4	5.5	36.7	63.4
Longitude	7.4	8.6	−9.4	23.3	6.2	7.7	−9.3	22.8
D near. city >= 10k	253	314	0	1701	109	133	0	1424
	Already existing cities >= 10k? at				Already existing city >= 10k? at			
0–20 km	0.05	0.22	0	1	0.09	0.29	0	1
20–50 km	0.19	0.39	0	1	0.33	0.47	0	1
50–100 km	0.40	0.49	0	1	0.58	0.49	0	1

Table A2
Sensitivity to the choice of city definition.

$P(city\ t \mid no\ city\ t-1)$	1 BASELINE >=5000	2 >=3000	3 >=4000	4 >=6000	5 >=7000	6 >=10,000	7 Step-wise
sea	0.007***	0.008***	0.008***	0.006***	0.006***	0.004***	0.005***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
river	0.012***	0.013***	0.013***	0.010***	0.008***	0.006***	0.008***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
hub	0.069***	0.075***	0.070***	0.057***	0.052***	0.033***	0.049***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
road	0.004***	0.005***	0.004***	0.003***	0.003***	0.002***	0.003***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
ln elevation	+0.00	+0.00	+0.00	+0.00	−0.00	−0.00	+0.00
	[0.98]	[0.69]	[0.69]	[0.74]	[0.79]	[0.72]	[0.82]
ruggedness	0.0001***	0.0001***	0.0001***	0.0001**	0.0001**	+0.00	+0.00
	[0.00]	[0.01]	[0.01]	[0.03]	[0.02]	[0.31]	[0.25]
P(cultivation)	0.0002**	0.0002***	0.0002**	0.0001*	0.0001	0.00	0.0001*
	[0.02]	[0.01]	[0.01]	[0.06]	[0.41]	[0.41]	[0.09]
city >= 10k? $(t-1)$							
0–20 km	−0.0005***	−0.0005***	−0.0005***	−0.0004***	−0.0002**	−0.0002*	−0.00036***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.02]	[0.05]	[0.00]
20–50 km	0.0002**	0.00014**	0.00014**	0.0001	0.0001	0.00004	0.00003
	[0.02]	[0.05]	[0.05]	[0.32]	[0.34]	[0.37]	[0.49]
50–100 km	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***	0.0001***	0.0001**
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.02]
$P(city\ t \mid no\ city\ t-1)$ unconditional	0.0005	0.0005	0.0005	0.0004	0.0003	0.0002	0.0003
NUTS2/century FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nr observations	2,588,903	2,588,401	2,588,655	2,589,204	2,589,388	2,589,720	2,589,090
p-value [20–100 km > 0?]	[0.00]	[0.00]	[0.00]	[0.02]	[0.01]	[0.02]	[0.08]
# NUTS regions	247	247	247	247	247	247	247
% w/ useful variation	23.6	26.2	24.5	20.4	18.5	13.2	17.9
% obs. in these NUTS	28.9	31.2	29.7	25.3	23.4	17.6	23.0

Notes: p-values, based on standard errors clustered at the NUTS3-century level between square brackets. +0.00 or −0.00 denotes a very small but positive, resp. negative, coefficient.
* Significance at the 10%.
** Significance at the 5%.
*** Significance at the 1%.

Again, only the effect of an already existing city within 20–50 km is sensitive to using this time-varying definition.[55]

A.3. 2nd nature geography results by construction? A Dartboard Approach

Given the steady increase in the number of cities over the centuries, one may be worried that especially our 2nd nature geography results could be obtained by construction. Is it simply this increased density of Europe's urban system that drives our finding of an increased importance of 2nd nature geography over the centuries?[56]

To assess this possibility we adopt the following *Dartboard Approach* in the spirit of Duranton and Overman (2005) and Ellison and Glaeser (1997). Using a simulation approach, we verify whether we would obtain the same results regarding our 2nd nature geography variables when cities appeared randomly at one of our potential locations instead of at the locations where they appeared in reality. If we do, this means we could be getting our results by construction, shedding doubts on our findings. This *Dartboard Approach* is operationalized as follows:

1. In each century t, we let cities appear randomly over the k_t available potential city locations in that century. We do this *conditional on each potential city location's 1st nature geography characteristics, i.e.:*

$$n_{it} \sim Binomial(1, p_{it}), \quad \text{where } p_{it} = X_i b + a_{rt}, \tag{B4}$$

   and b an a_{rt} are the estimated parameters on the 1st nature geography variables and the estimated NUTS2-century fixed effects respectively, obtained by estimating (2) including only these 1st nature geography variables and NUTS2-century dummies as explanatory variables.
2. Using this hypothetical city configuration, we estimate our baseline model as in (2) and store the estimated parameters on each of the three 2nd nature geography variables.
3. Repeat the above-outlined procedure 1000 times. Next, for each respective 2nd nature geography variable we calculate the percentage of simulation runs that we *falsely do not reject* that it is significant at the 1%, 5%, and 10% respectively when using standard z-tests. Given that we allocate new cities randomly in each century, this percentage should be close to 1%, 5%, and 10% respectively to conclude that the standard tests perform well.

Table A3 shows the results of doing these simulations for our main specification (column 1 in Table 1). The percentage of simulation runs in which we falsely reject the null hypothesis is always quite close to 1%, 5% or 10% respectively, providing strong confidence that we do not obtain results by construction as a consequence of the increased density of the urban system over the centuries.

[55] Results are also robust to using the city definition in period t instead. Also, using a different 'step-wise' city definition (i.e. 5000 before 1800 and 10,000 in 1800) all our baseline results come through. These results are available upon request.

[56] Note that this issue is related, to the possibility of dynamic selection bias. However, where dynamic selection bias concerns the dependent variable, the concern that we address here is that the increasing number of cities over time affects our 2nd nature geography regressors with possibly unwanted consequences for our results.

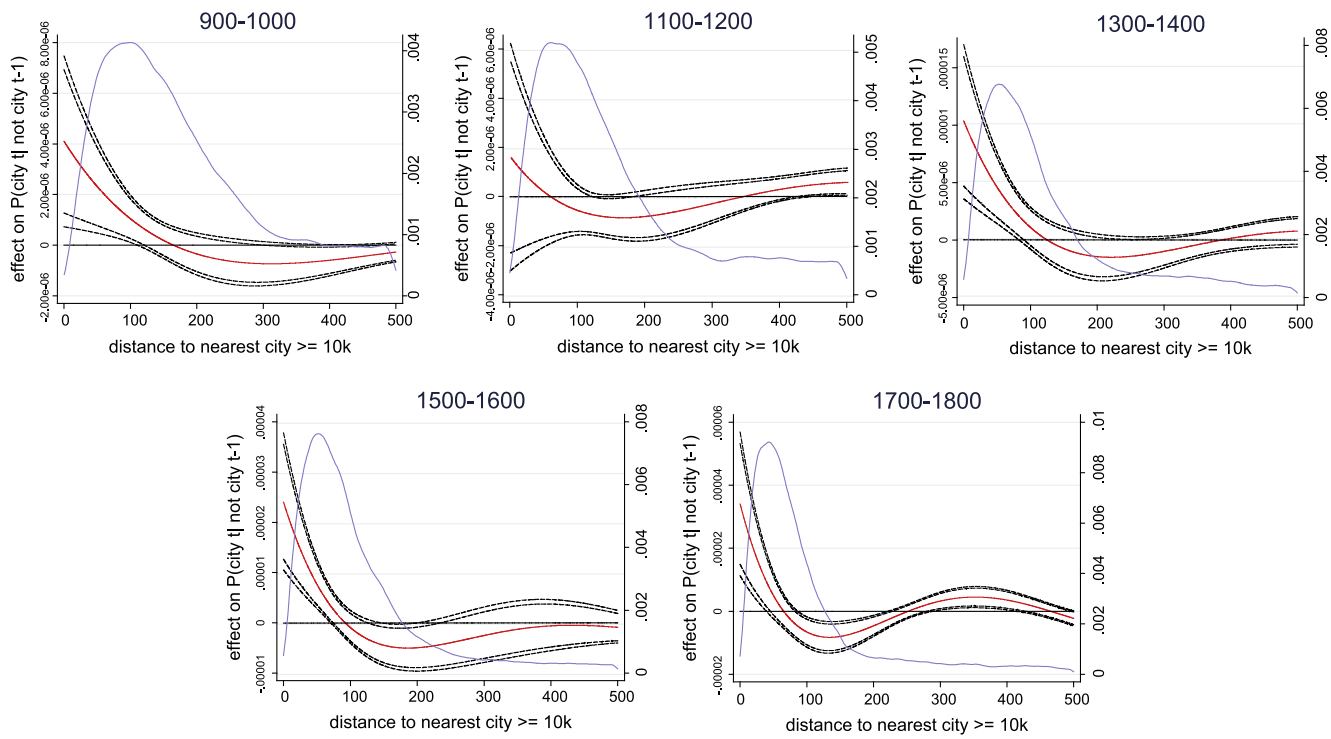


Fig. A3. More elaborated 2nd nature geography over space and time [estimating (3) by 200 year intervals]. *Notes:* distances are in kilometers. The gray line in each figure depicts the distribution of distance to the nearest city over all locations in our sample in each specific 200-year period. See Table A4 for more info on these distributions. The dotted lines show 5% and 10% confidence intervals around the estimated marginal effect at each distance to the nearest city. They are calculated using standard errors clustered at the NUTS3/century level. Results for 1100–1200 should be taken with some care (see the discussion in the main text above Table 3).

Table A3
Dartboard Approach – simulation results.

	% falsely not rejected at		
	1%	5%	10%
city >= 10k? $(t-1)$			
0–20 km	1.1	5.2	11.4
20–50 km	1.1	6.7	11.6
50–100 km	1.0	4.2	9.7

Notes: All the '% falsely not rejected' are based on 1000 simulation runs. In all simulation runs standard errors are clustered at the NUTS3/century level.

Because of the ever denser urban system the number of locations *becoming a city*, and located close(r) to an already existing city, increases over the centuries; but, so does the number of locations *not becoming a city*, located close(r) to an already existing city. The estimated parameters on our 2nd nature geography dummy variables depend on the trade-off between these two. Our simulation results show that their significance is not simply an artefact of the increased density of the European urban system in the later centuries.

A.4. A spatial and temporal extension of our 2nd nature geography results

In this section we show results when estimating (3), while also allowing the polynomial in distance to the nearest already existing city to change over the centuries. Fig. A3 shows how the distance to the nearest already existing city affects a location's urban chances in each of the five 200-year periods in which we divide

Table A4
Accompanying info to Fig. A3.

Period	Distance range urban shadow		Point estimate turns negative at	Distance range positive 2nd nature effect	
	Sign. 5%	Sign. 10%		Sign. 5%	Sign. 10%
900–1000	<111 km	<121 km	165 km	–	301–475 km
1100–1200	–	–	61 km	–	127–171 km
1300–1400	<82 km	<89 km	122 km	–	229–282 km
1500–1600	<71 km	<75 km	94 km	153–193 km	130–228 km
1700–1800	<39 km	<45 km	65 km	86–222 km	82–228 km
	5%	25%	50%	75%	95%
Quantiles of the distance to nearest already-existing city distribution					
900–1000	43 km	120 km	262 km	693 km	1297 km
1100–1200	30 km	82 km	161 km	472 km	1099 km
1300–1400	22 km	61 km	113 km	312 km	984 km
1500–1600	21 km	54 km	97 km	204 km	719 km
1700–1800	16 km	41 km	72 km	133 km	571 km

Notes: All distances correspond to the results and distributions shown in Fig. A3.

our sample period. Table A4 provides accompanying information on the reach of an existing city's urban shadow as well as the distance range at which an existing city positively affects other locations' urban chances (again for each of these five 200-year periods).

Appendix B. Theory appendix

In this Appendix, we briefly set out the main features of the economic geography models (Fujita and Krugman, 1995; Fujita and Mori, 1997; Fujita et al., 1999) whose predictions we take to the data in our paper. We do this by means of discussing one particular model in more detail, that of Fujita and Mori (1997). Our discussion and notation closely follow that in Fujita and Mori (1997). Note that the theoretical exposition in this Appendix is only meant to *illustrate* the main features of this type of economic geography models. There are other observationally equivalent ways to microfound the trade-off between agglomeration and dispersion forces that lie at the heart of these models (e.g. production linkages, knowledge spillovers, easier matching on the labor market, or congestion).

We focus on the predictions of the model as to how an already-existing city influences the urban chances of its immediate surroundings. We mention the main forces in the model that determine whether or not a new city can emerge, and if so, how the exact location of its emergence depends on its distance to the already existing city. For the complete model setup, we refer to the original paper.

Consider an economy that can be represented by a one-dimensional unbounded space, X. To abstract from any (1st nature geography) differences between locations, the quality of land is homogenous and its density is 1 everywhere (1 unit of land per unit of distance). The economy consists of two sectors, agriculture (A) and manufacturing (M). Agricultural production takes place under a Leontief technology requiring a_A units of labor and one unit of land to produce one unit of output. The manufacturing sector instead supplies a continuum of differentiated varieties, whose production takes place under an increasing returns to scale technology using labor only. To produce Q units of any variety, ω, of the M-good requires the same $L = f + a_M Q$ units of labor, where f and a_M are the fixed and marginal input requirements. Because of these scale economies in production, each variety of the M-good is produced by a single firm.

The economy consists of N workers, each endowed with 1 unit of labor, and landlords. Landlords are attached to their land consuming all revenue from their land at their location. Workers instead are freely mobile between locations and sectors. Both landlords and workers have the same utility function:

$$U = z_A^{\alpha_A} \left[\int_0^n z_M(\omega)^\rho d\omega \right]^{\alpha_M/\rho}, \tag{B1}$$

where z_A denotes the amount of the A-good consumed, $z_M(\omega)$ the consumption of each manufacturing variety $\omega \in [0,n]$, α_A and α_M the expenditure shares on the A-good and the M-goods respectively such that $\alpha_A + \alpha_M = 1$, and ρ reflects consumer's preference for variety in the M-goods produced (the lower ρ, the more differentiated manufacturing varieties are from a consumer's perspective).

Location crucially matters in this model because selling goods produced in one location, $r \in X$, in another location, $s \in X$, incurs transportation costs that take the standard Samuelson's iceberg form, i.e. a fraction $e^{-\tau_i|s-r|}$ of the goods produced in r arrive in s, where τ_i is a constant that can differ between the A- and the M-sector, i.e. $i \in (A,M)$.

Consumers in each location maximize their utility subject to their budget constraint and a given set of prices. This pinpoints demand for each M-variety, and for the A-good, produced in location $r \in X$ from consumers in all other locations $s \in X$. Similarly, if a firm chooses to locate in location r, it sets its price in order to maximize its profits given demand from each other location s, given wages and given transport costs (it also takes other firms' prices as given, i.e. firms behave in a non-strategic manner: a result of the monopolistic competitive nature of the M-sector). This in turn pinpoints supply of each M-variety, and of the A-good, produced in location $r \in X$ in all other locations $s \in X$.

In equilibrium, goods and labor markets clear, and profits go to zero in both the M- and the A-sector (free entry of firms). Moreover, workers move between locations until real wages are equalized across locations. This determines the prevailing wages and prices in all locations, and, most importantly for our purpose, the spatial distribution of workers and firms across all locations $x \in X$ (implicitly also defining the sectoral composition in each location).

Generally however, multiple equilibrium solutions exist, each with a different spatial distribution of firms and workers (see Fujita and Mori, 1997; or Fujita and Krugman, 1995). In principle this would make it very difficult to draw clear-cut conclusions. However, to overcome this issue, all papers propose an approach that allows them to focus on the following interesting questions. First, they consider the existence of only one city (i.e. all M-sector activity located in a single location), and ask *under what conditions such a monocentric city configuration is an equilibrium?* Second, they ask whether the monocentric equilibrium is unstable, i.e. *under what conditions would it be viable for workers and firms to found a second city*? Moreover, *where would the new city be located*?

The answers to these questions are the theoretical underpinnings of the empirical exercise in our paper. They depend on the strength of two opposing forces: *agglomeration* and *dispersion* forces. Agglomeration forces arise through the interaction of product-variety in M-goods and the scale economies in their production. Against them go the dispersion forces induced by the demand for manufactured consumption goods by the dispersed agricultural population, and the demand for agricultural goods by the city population. Transportation costs play a crucial role in determining the relative strength of these forces.

Table B1 shows the parameter conditions under which the monocentric city configuration is a possible equilibrium outcome (see also p. 515 in Fujita and Krugman (1995) or p. 412 in Fujita and Mori (1997)):

The parameters determining the stability of the monocentric equilibrium are: consumers' expenditure shares on agricultural and manufacturing products (α_A and α_M), transport costs in both sectors (τ_A and τ_M), and ρ determining consumers' preference for variety (or alternatively, the extent of scale economies in manufacturing production). Three cases can be distinguished.

Case 1: the monocentric city configuration is never a stable equilibrium. When consumers primarily spend their income on agricultural products, and transport costs for agricultural products are much larger than those for manufacturing goods (see condition (*ii*)), no cities will ever develop. In this case, since the supply of agricultural goods is tied to the land, consumers choose to locate in the hinterland instead of agglomerating in a city. Clearly, this particular case is *empirically irrelevant: we do observe cities in Europe throughout our sample period.*[57]

That makes case 2 and 3 the more interesting, the monocentric city configuration is a spatial equilibrium in these two scenarios:

[57] Note that in case of heterogeneity in the main model parameters across Europe, this case could be relevant. Some regions (e.g. Scandinavia, Poland, Austria or Ireland) do not contain any cities in the beginning of our sample period. The first emergence of a city in these parts during our sample period, is consistent with falling agricultural transport costs and/or a decrease in the share of income spent on agricultural goods in these regions.

Table B1
Stable monocentric city equilibrium?

(i) $\alpha_A\tau_A \le (1+\rho)\alpha_M\tau_M$		(ii) $\alpha_A\tau_A > (1+\rho)\alpha_M\tau_M$
(iii) $\alpha_M \ge \rho$	(iv) $\alpha_M < \rho$	(1) Never
(2) Always	**(3) Depends**	

cities emerge. When consumers spend more of their income on manufacturing goods, and the costs involved in transporting food into the city are relatively low, the benefits of co-locating in the city (due to the scale economies in the production of M-varieties) outweigh the costs involved in shipping agricultural products from the hinterland to the city. As a result, a spatial configuration characterized by at least one city is an equilibrium outcome in both case 2 and 3.

Case 2: the lure of the existing city always prevents new cities from appearing anywhere in the hinterland. This happens when in addition to satisfying condition (i), manufacturing varieties are relatively more differentiated, and consumers spend a large share of their budget on these varieties (condition (iii) is satisfied). Also this scenario is clearly *empirically irrelevant*: *we do observe new cities emerging all over Europe in each and every century in our sample.* This leaves us case 3, as the only empirically relevant scenario (Fujita and Mori (1997) say that case 1 and 2 "*are not interesting cases for our study of the evolution of an urban system*" [p. 412]).

Case 3: if manufacturing varieties are less differentiated (the scale economies not too strong), and/or consumers spend a smaller share of their income on manufacturing goods (condition (iv) is satisfied), the monocentric city equilibrium is not necessarily stable: *the possibility for a new city to emerge arises*. The natural question to ask in this, empirically most relevant, case then is: when does a new city emerge and where will it be located?

In order to answer these questions Fujita and Krugman (1995) and Fujita and Mori (1997) proceed as follows: suppose the initial city is located at the origin, *0*. In order for this to be a spatial equilibrium, no firm should be able to make a profit by moving to any other location, $x \in X$. The authors show (see e.g. Eq. (22) in Fujita and Mori (1997)) that for each location, $x \in X$, its so-called *market potential*, Ω, can be calculated as a function of the model's main parameters, ρ, α_A, α_M, τ_A, and τ_M, its distance to the existing city, x, and overall population, N:

$$\Omega(x;N) = e^{-\eta x}\left\{1 + \alpha_A\gamma\tau_M\int_0^x e^{2\gamma\tau_M y}\left(1 - \frac{1-e^{-\tau_A y}}{1-e^{-\tau_A l^*(N)}}\right)dy\right\}, \tag{B2}$$

where $\gamma = \rho/(1-\rho)$, $\eta = (1+\gamma)(\alpha_M\tau_M - \alpha_A\tau_A) + \gamma\tau_M$, and $l^*(N)$ denotes the agricultural boundary of the city,[58] which in equilibrium is uniquely determined by an increasing function of population size, N (the more people, the more agricultural land is required to feed them; see e.g. Eq. (15) in Fujita and Mori, 1997). If a location's market potential is above unity, a firm that decides to move away from the existing city to that location will make a profit. Other firms will follow, and a new city emerges.

Calculating (B2) at different distances from the existing city thus reveals where a new city can in principle develop. Fig. B1, depict these market potential curves for the same parameter setting as the one used in Fig. 3 in Fujita and Mori (1997). It shows, that population plays a crucial role in determining whether or not new cities emerge. When population is too small, no new cities emerge. However, as population grows, more and more locations become viable new city locations, i.e. the larger the set of $x \in X$ for which $\Omega(x,N) > 1$. In fact, from (B2) and the fact that $l^*(N)$ is an increasing

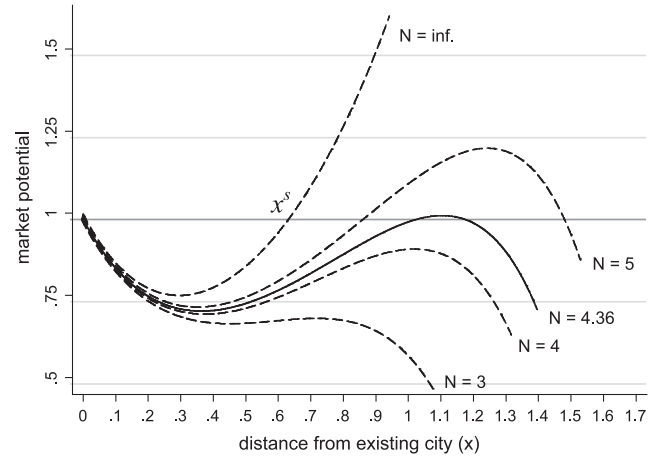


Fig. B1. Market potential curves – Fig. 3 in Fujita and Mori (1997). *Notes:* the different curves are calculated keeping all other parameters fixed at: $\rho = 0.75$, $\alpha_M = \alpha_A = 0.5$, $\tau_M = 1$, $\tau_A = 0.8$, $f = 1$, $a_M = 1$, and $a_A = 0.5$. N is total population.

function of N, it can be easily seen that: $\partial\Omega(x;N)/\partial N > 0\ \forall x > 0$: a growing population unambiguously increases all locations' market potential.

Fig. B1 also shows that whether or not *any particular location $x \in X$, becomes a viable new city location depends first and foremost on its distance to an already existing urban center*. Locations too close to an already existing city face too strong competition with that city, both for agricultural produce and for inhabitants. On the other hand, locations too far from an already existing city cannot take full advantage of the trading possibilities with the already existing city. This leaves *locations at medium range from existing cities as preferred new city locations*: they offer relatively cheap trading possibilities with the already existing cities compared to locations further off (i.e. better market access), as well as only limited competition with these same existing cities compared to locations at too close range.

As population continues to grow, the range of locations that become viable new city locations expands (both toward locations further away from, as well as closer to the already existing city). However, with an ever expanding population, locations too close to the already-existing city will never have any chance of developing into a city themselves: the existing city exerts a strong urban shadow that will always prevent the formation of new cities in its immediate backyard.

Total population is (of course) not the only variable affecting the range of viable new city locations. However, establishing how the other main model parameters affect this range is much more difficult. Contrary to total population, a change in ρ, α_A, α_M, τ_A, or τ_M, does not unambiguously in- or decrease market potential for each location $x > 0$. This depends on the other model parameters as well as total population; moreover it can differ at different distances, x, to the existing city.

One property can however be established. This is how ρ, α_A, α_M, τ_A, or τ_M affect the minimum reach of the existing city's urban shadow: the maximum distance from the existing city where a new city will never become viable no matter how much total population increases [x^s in Fig. B1]. Fujita and Mori (1997, p. 411) show that this reach can be represented by the absolute value of the slope of the market potential curve (B2) at $x = 0$:

$$\theta = [(1+\rho)\alpha_M\tau_M - \alpha_A\tau_A]/(1-\rho). \tag{B3}$$

From (B3), one can immediately see that the reach of the existing city's urban shadow increases with ρ, α_M and τ_M, and decreases with α_A and τ_A (given that their change does not result in a violation of condition (i) or (iv) in Table B1).

[58] Beyond this agricultural boundary, it is not profitable to produce agricultural goods and ship them to the city.

Summing up, in the empirically most relevant case 3, new cities can emerge. Their emergence becomes more likely, the larger total population. Moreover, they do not emerge everywhere, locations at medium distance to an existing city are preferred city locations. We take these predictions to the data in our paper.

References

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

Anas, Alex, Arnold, Richard, Small, Kenneth A., 1998. Urban spatial structure. Journal of Economic Literature 26, 1426–1464.

Ashton, Thomas, 1948. The Industrial Revolution (1760–1830). Oxford University Press, Oxford.

Bahn, Paul, 1996. Archaeology: A Very Short Introduction. Oxford University Press, Oxford.

Bairoch, Paul, 1988. Cities and Economic Development. From the Dawn of History to the Present. University of Chicago Press, Chicago.

Bairoch, Paul, Batou, Jean, Pierre, Chèvre, 1988. La population des villes Européennes, 800–1850. Librairie Droz, Genève.

Behrens, Kristian, 2007. On the location and lock-in of cities: geography vs. transportation technology. Regional Science and Urban Economics 37, 22–45.

Black, Duncan, Henderson, Vernon, 1999. A theory of urban growth. Journal of Political Economy 107 (2), 252–284.

Black, Duncan, Henderson, Vernon, 2003. Urban evolution in the USA. Journal of Economic Geography 3 (4), 343–372.

Bleakley, Hoyt, Lin, Jeffrey, 2010. Portage: Path Dependence and Increasing Returns in U.S. History. Working Paper, Federal Reserve Bank of Philadelphia.

Bosker, Maarten, Buringh, Eltjo, van Zanden, Jan Luiten, 2013. From Baghdad to London: Unraveling urban development in Europe, the Middle East, and North Africa, 800–1800. Rev. Econ. Stat. 95 (4), 1418–1437.

Bosker, Maarten, Buringh, Eltjo, 2013. City Seeds: Geography and The Origins of The European City System. Working Paper, Erasmus University Rotterdam.

Christaller, Walter, 1935. Die zentralen Orte in Suddeutschland. Jena.

Davis, Kingsley, 1955. The origin and growth of urbanization in the world. American Journal of Sociology 60, 429–437.

Davis, Donald, Weinstein, David, 2002. Bones, bombs and breakpoints: the geography of economic activity. American Economic Review 92, 1269–1289.

DeLong, J. Bradford, Shleifer, Andrei, 1993. Princes and merchants: city growth before the industrial revolution. Journal of Law and Economics 36, 671–702.

Duranton, Gilles, 1999. Distance, land, and proximity: economic analysis and the evolution of cities. Environment and Planning A 31, 2169–2188.

Duranton, Gilles, 2007. Urban evolutions: The fast, the slow, and the still. Am. Econ. Rev. 197–221.

Duranton, Gilles, Overman, Henry, 2005. Testing for localization using micro-geographic data. Review of Economic Studies 72, 1077–1106.

Dyer, Chris, 1995. How urbanized was medieval England?. In: Duvosquel, J.M., Thoen, E. (Eds.) Peasants and Townsmen in medieval Europe. Snoek Ducaju & Zoon, Gent, pp. 169–183.

Eeckhout, Jan, 2004. Gibrat's law for (all) cities. Am. Econ. Rev. 1429–1451.

Ellison, Glenn, Glaeser, Edward E., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. Journal of Political Economy 105 (5), 889–927.

Ennen, Edith, 1972. Die Europäische Stadt des Mittelalters. Vandenhoeck und Ruprecht, Göttingen.

Fagan, Brian M., 1997. Archaeology: A Brief Introduction. Longman, New York.

Findlay, Ronald, O'Rourke, Kevin H., 2007. Power and Plenty: Trade, War, and the World Economy in the Second Millennium. Princeton University Press, Princeton.

Fujita, Masahisa, Krugman, Paul, 1995. When is the economy monocentric? Regional Science and Urban Economics 25, 505–552.

Fujita, Masahisa, Krugman, Paul, Mori, Tomoya, 1999. On the evolution of hierarchical urban systems. European Economic Review 43 (2), 209–251.

Fujita, Masahisa, Mori, Tomoya, 1996. The role of ports in the making of major cities: self-agglomeration and hub-effect. Journal of Development Economics 49, 93–120.

Fujita, Masahisa, Mori, Tomoya, 1997. Structural stability and evolution of urban systems. Regional Science and Urban Economics 27 (4–5), 399–442.

Galor, Oded, 2005. From stagnation to growth: unified growth theory. In: Aghion, Phillipe, Durlauf, Steven N. (Eds.), Handbook of Economic Growth. Elsevier, Amsterdam.

Glaeser, Edward L., 2011. The Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier and Happier. Macmillan, New York.

Greif, Avner, 1992. Institutions and international trade: lessons from the commercial revolution. American Economic Review – Papers and Proceedings 82 (2), 128–133.

Greif, Avner, 2006. Institutions and the Path to the Modern Economy: Lessons from Medieval Trade. Cambridge University Press.

Hajnal, J., 1965. European marriage in perspective. In: Glass, David V., Eversley, David E.C. (Eds.), Population in History. Arnold, London, pp. 101–143.

Hanson, Gordon, 2005. Market potential, increasing returns and geographic concentration. Journal of International Economics 67 (1), 1–24.

Hausman, Jerry, 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. Journal of Economic Perspectives 15 (4), 57–67.

Henderson, Vernon, 1974. The sizes and types of cities. American Economic Review 64, 640–656.

Hohenberg, Paul M., Lees, Lynn H., 1995. The Making of Urban Europe: 1000–1994. Harvard University Press, Cambridge.

Hybel, Nils, 2002. The grain trade in northern Europe before 1350. Economic History Review 55 (2), 219–427.

Konishi, Hideo, 2000. Formation of hub cities: transportation cost advantage and population agglomeration. Journal of Urban Economics 48 (1), 1–28.

Krugman, Paul, 1991. Increasing returns and economic geography. Journal of Political Economy 99, 483–499.

Krugman, Paul, 1993. The hub effect: or, threeness in interregional trade. In: Ethier, Wilfred J., Helpman, Elhanan, Neary, J. Peter (Eds.), Theory, Policy and Dynamics in International Trade. Cambridge University Press, Cambridge, pp. 29–37.

Krugman, Paul, 1993. On the number and location of cities. European Economic Review 37 (2–3), 293–298.

Lampard, Eric, 1955. The history of cities in the economically advanced areas. Economic Development and Cultural Change 3 (2), 81–136.

Lopez, R.S., 1956. The evolution of land transport in the middle ages. Past and Present 9, 17–29.

Lösch, August, 1938. The nature of economic regions. Southern Economic Journal 5, 71–78.

Lösch, August, 1940. Die räumliche Ordnung der Wirtschaft (Gustav Fischer, Jena); English translation: The Economics of Location (Yale University Press, New Haven, 1954).

Malanima, Paolo, 1998. Italian Cities 1300–1800. A quantitative approach. Rivista di Storia Economica 91–126.

Marshall, Alfred, 1890. Principles of Economics, eighth ed. Macmillan, London.

McEvedy, Colin, Jones, Richard, 1979. Atlas of World Population History. Allen Lane, London.

Nunn, Nathan, Qian, Nancy, 2011. The potato's contribution to population and urbanization: evidence from a historical experiment. Quarterly Journal of Economics 126, 593–650.

Pirenne, Henri, 1925. Medieval Cities, Their Origins and The Revival of Trade. Princeton University Press, Princeton.

Redding, Stephen, Sturm, Daniel, 2008. The costs of remoteness: evidence from German division and reunification. American Economic Review 98 (5), 1766–1797.

Ramankutty, N., Foley, J.A., Norman, J., McSweeney, K., 2002. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. Global Ecology and Biogeography 11 (5), 377–392.

Stewart, John Q., 1947. Empirical mathematical rules concerning the distribution and equilibrium of population. Geographical Review 37, 461–485.

Stewart, Mark, 2007. The inter-related dynamics of unemployment and low-wage unemployment. Journal of Applied Econometrics 22, 511–531.

Stewart, Mark, 2006. Maximum simulated likelihood estimation of random effects dynamic probit models with autocorrelated errors. Stata Journal 6 (2), 256–272.

Talbert, Richard J.A. (Ed.), 2000, Barrington Atlas of the Greek and Roman World. Princeton University Press, Princeton.

Von Thünen, Johann Heinrich, 1826. Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalokonomie. Perthes, Hamburg.

Ullman, Eric, 1941. A theory of location for cities. American Journal of Sociology 46 (6), 853–864.

de Vries, Jan, 1984. European Urbanization, 1500–1800. Methuen, London.

Weber, Max, 1922. The City (Translation and edited by Don Martindale and Gertrud Neuwirth). The Free Press, New York.

Woolridge, Jeffrey, 2005. Simple solutions for the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics 20, 39–54.

van Zanden, Jan Luiten, van Tielhof, Milja, 2009. Roles of growth and productivity change in Dutch shipping industry, 1500–1800. Explorations in Economic History 46 (4), 389–403.