University of Glasgow

Adam Smith Business School

WORKING PAPER SERIES

The Death and Life of Great British Cities

Stephan Heblich, Dávid Krisztián Nagy, Alex Trew and Yanos Zylberberg

Paper No. 2023-09
July 2023

The Death and Life of Great British Cities*

Stephan Heblich Dávid Krisztián Nagy Alex Trew
Yanos Zylberberg

July 3, 2023

Abstract

This paper studies how cities' industrial structure shapes their life and death. Our analysis exploits the large heterogeneity in the early composition of English and Welsh cities. We extract built-up clusters from early historical maps, identify settlements at the onset of the nineteenth century, and isolate exogenous variation in the nature of their rise during the transformation of the economy by the end of the nineteenth century. We then estimate the causal impact of cities' population and industrial specialization on their later dynamics. We find that cities specializing in a small number of industries decline in the long run. We develop a dynamic spatial model of cities to isolate the forces which govern their life and death. Intratemporally, the model captures the role of amenities, land, local productivity and trade in explaining the distribution of economic activity across industries and cities. Intertemporally, the model can disentangle the role of aggregate industry dynamics from city-specific externalities. We find that the long-run dynamics of English and Welsh cities is explained to a large extent by such dynamic externalities à la Jacobs.

Keywords: specialization; cities over time; quantitative economic geography.
JEL codes: F63, N93, O14, R13

*Heblich: University of Toronto, CESifo, IfW Kiel, IZA, SERC; stephan.heblich@utoronto.ca; Nagy: CREi, Universitat Pompeu Fabra and BSE; dnagy@crei.cat; Trew: University of Glasgow; alex.trew@glasgow.ac.uk; Zylberberg: University of Bristol, CEPR; yanos.zylberberg@bristol.ac.uk. We thank Fabian Eckert, Javier Quintana and Lin Tian, as well as participants in seminars and conferences, e.g., at Boston University, Chicago Fed, CRED, CREI, CURE, FREIT, McGill, Philadelphia Fed, Princeton, Richmond Fed, SED, UEA, Universitat Pompeu Fabra, University of Bologna, University of Oslo, University of Southern California, University of Virginia and Warwick for very useful comments and suggestions. We acknowledge support of the ANR/ESRC/SSHRC, through the ORA grant ES/V013602/1 (MAPHIS: Mapping History).

Many of the great cities that drove the industrial transformation of the nineteenth century have declined in the twentieth century. The formerly thriving mining towns of south Wales, the rust belt cities in the northeastern United States, or the Ruhr valley in Germany once employed generations of workers but now struggle to find renewed economic success. Both the initial rise of those cities and the possible reversal of fortune may be tied to the dynamics of their industries. First, there are external, *macroeconomic* factors driving the aggregate dynamics of industries within a country, and the dynamics of cities might be tied to the aggregate dynamics of industries in which they have a comparative advantage. Second, *internal* factors might also matter for how a city's industrial portfolio affects its long-run development. For instance, Glaeser et al. (1992) discuss the role of dynamic, between-industry "Jacobs" externalities in driving city productivity—as motivated by the seminal work of Jacobs (1969) describing the "Death and Life of Great American Cities".

This paper studies how cities' local industrial structure shapes their life and death. We rely on unique data characterizing the spatial distribution of industries in England and Wales over the course of two centuries. More specifically, (i) we observe the settlements of the early nineteenth century in historical maps, which allows us to isolate clusters of built-up areas to delineate the locations of cities; (ii) we can follow their population and industrial composition using censuses (1801–1911) and a quasi-census based on baptism records (1817) during the rapid industrialization of Britain in the nineteenth century; and (iii) we can measure their long-run economic performance using an array of contemporary high-quality data. This helps us characterize the death and life of Great British cities.

Britain is an ideal setting for our analysis. Its economic geography was transformed by the invention of new technologies in key sectors in the eighteenth century, which led to rapid industrialization and urbanization in the nineteenth century after the conclusion of wars against France. Rapid growth in technology and international trade, and the resulting shift to large-scale, steam-powered industrial production, transformed both the scale and location of industries that underpinned the sustained increase in aggregate growth. Alongside this growth, the mostly small settlements of the early nineteenth century gave way to a dense network of cities by the end of the nineteenth century. Some cities specialized in a few industries; others were more diverse. The large heterogeneity in population and the industrial composition of English and Welsh cities at that time induced contrasting local dynamics during the twentieth century.

To estimate the causal impact of cities' population and specialization on their long-run dynamics, we need to isolate observable, exogenous variation in the nature of their rise during the industrialization wave of the nineteenth century. We do so in two steps.

In a first step, we collect unique historical county maps from around 1790–1820, develop a machine-learning algorithm to identify built-up areas, and select urban settlements as sufficiently large clusters of built-up areas. The output of this procedure is a set of settlements—with their boundaries in the early nineteenth century—that have the potential to develop into cities by the late nineteenth century. In a second step, we construct two exogenous predictors of a city's economic structure at the end of the nineteenth century: a predictor for city population; and a predictor for city specialization.

To predict late-nineteenth century city *population*, we borrow insights from the historical literature and conjecture that the fragmentation of land ownership in the immediate fringe of cities affected the pace at which they could grow in response to industrialization. The underlying argument is at the heart of the land assembly problem (see, e.g., Eckart, 1985; Strange, 1995): a higher degree of land ownership fragmentation in the city's fringe makes negotiations to acquire additional land for urban use more costly. To isolate exogenous variation in land fragmentation over cities' initial fringe (1790–1820), we exploit plausibly exogenous breaks in terrain characteristics including elevation, ruggedness, persistent soil attributes and water bodies (controlling for the *average* ruggedness around city boundaries, see Saiz, 2010; Harari, 2020). We develop an algorithm to predict the natural delineation of fields from multidimensional breaks in these natural characteristics.[1] One may think of this algorithm as identifying natural fault lines between potential agricultural land parcels. Importantly, while the local characteristics of city fringes at the beginning of the nineteenth century did influence city growth in the following decades, they did not directly affect the later evolution of cities during the twentieth century. This is because all cities grew out of this narrow ring and were thus subject to another topography at their now expanded borders. To predict the potential for city *specialization*, we rely on a measure of initial location advantages—the industrial composition of early settlements in 1817—and create a "shift-share Herfindahl index" that combines initial sectoral shares with the aggregate employment growth of sectors. By the end of the nineteenth century, cities of England and Wales have markedly different economic structures: some cities have grown large, while others have not; some cities have developed a diverse industrial base, while others have specialized in a narrow set of industries.

We find that cities which specialize in a small number of industries at the end of the nineteenth century experience a decline in the long run. By contrast, we do not find city size at that time to be a significant predictor of future growth—consistent with

[1] We validate the predictive power of this measure of land fragmentation by comparing it with the actual observed concentration of ownership from land tax registers, Tithe records and later micro-census records where land acreage is reported by landowners.

Gibrat's law. We subject these results to an array of robustness checks that include controlling for agricultural productivity, agricultural mechanization, and differences in institutions related to land ownership. Crucially, the results are also robust to controlling for aggregate industry dynamics captured by the nationwide employment growth of industries, both during the nineteenth and during the twentieth century.

To interpret these stylized facts, we develop a multi-sector dynamic spatial model and study the mechanisms behind the evolution of industries across cities and over time. The model features a finite set of cities that trade the products of a finite set of industries. At any point in time, cities can differ in their sectoral productivities, their amenities, their land supply elasticity, and their trade costs with other cities. Population mobility is subject to frictions across cities. City population and industrial structure can influence future productivity in a flexible way. In particular, the model allows for both dynamic Marshall–Romer and Jacobs externalities, which have been discussed as important determinants of city growth and industrial structure in the literature (see Carlino and Kerr, 2015, for a review).[2]

The model offers the following rationale behind (i) the distribution of economic activity at the end of the nineteenth century and (ii) the subsequent dynamics. By the end of the nineteenth century, trade costs are low, leading to a specialization of cities in their comparative advantage sectors. This increased specialization is particularly pronounced in the most centrally-located cities, which experience the largest gains from trade. The availability of land (as captured by the city-specific land supply elasticity) disciplines the degree to which population reallocates towards these cities. With dynamic externalities and aggregate industry trends, this initial distribution of employment across industries and cities leads to markedly different city dynamics across space. First, industries that were successful earlier may turn into a period of decline over their life-cycle, e.g., because of an aggregate structural change of the economy (Ngai and Pissarides, 2007) or because of international competition and exposure to trade (Pierce and Schott, 2016). The model then shows how cities with a location advantage in floundering industries appear to suffer from their initial specialization. Second, dynamic Jacobs externalities, generating larger productivity gains in historically more diverse cities, might direct population and economic activity away from cities that were highly specialized in the past, irrespective of aggregate trends in these areas of specialization. We illustrate the model's ability to disentangle these two mechanisms by simulating the model on a simple (linear) geography. In the last part of the paper, we offer a strategy of taking the full model to

[2]Marshall–Romer externalities operate within industry, implying that cities specialized in a narrow set of industries are the ones primarily benefiting from these externalities. By contrast, Jacobs externalities operate across industries and hence favor cities with a diverse industrial base.

the data to conduct this decomposition.

We contribute to several strands of existing research. First, we relate to the literature on industries as drivers of urban growth (Duranton, 2007; Hanlon and Miscio, 2017), and specifically to contributions that discuss the negative effects of specialization on city development (Glaeser et al., 1992; Duranton and Puga, 2001; Faggio et al., 2017; Heblich et al., 2019). One feature of our model, responsible for the negative effect of specialization, is the crucial role played by externalities à la Jacobs (1969) as drivers of development in the long run (Carlino and Kerr, 2015). The closest paper to ours in this literature is Henderson et al. (1995), which discusses the life-cycle of industries and cities. In a more specific application, Glaeser (2005) studies Boston over nearly four centuries and points to the important role of human capital in reinventing Boston after periods of crisis and decline. Alternative mechanisms behind the rise and fall of cities could be distortion in the acquisition of human capital (see, e.g., Franck and Galor, 2021) or the life cycle of industries (Henderson et al., 1995). The key advantage of our empirical strategy is that it allows us to quantify the long-run causal impact of specialization on city outcomes, while accounting for the life cycle of industries.

Second, we connect to a quantitative literature studying the dynamic evolution of the spatial distribution of economic activity. The closest contributions in this literature are Allen and Donaldson (2020), Berkes et al. (2021), Caliendo et al. (2019), Nagy (2023), Eckert and Peters (2023) and Fajgelbaum and Redding (2021).[3] Our main contribution to this literature lies in proposing a multi-sector dynamic model with various dimensions of heterogeneity (trade costs, sectoral productivities, amenities and land supply elasticities) that can be taken to the data in a computationally tractable way.

Third, we contribute to the literature on the drivers of the industrial revolution (for a survey, see Clark, 2014). Stokey (2001) showed the quantitative importance of trade for structural transformation at a macroeconomic level, while Allen (2009) has argued that the mechanism at work is via the impact of trade on relative prices. We suggest another potential channel, i.e., that trade induced the growth of manufacturing in cities, accelerating the transition to large-scale, export-oriented growth in the nineteenth century. Lastly, the focus on the first structural transformation *and* the second transition away from industry in the twentieth century also links to the macroeconomic literature on growth and structural transformation summarized in Herrendorf et al. (2014).

The remainder of the paper is organized as follows. Section 1 presents some historical context. Section 2 describes our data sources, data construction, and empirical strategy.

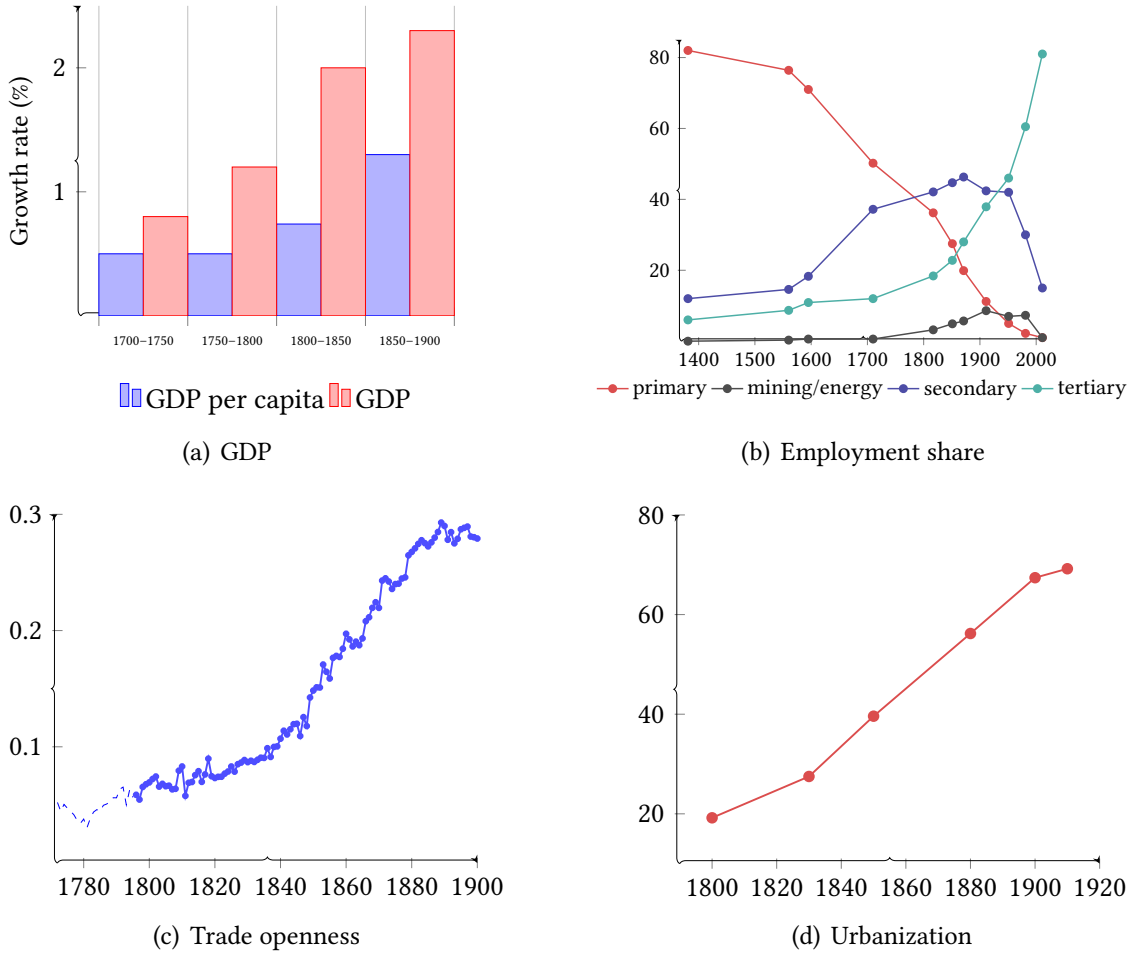
[3]For a more comprehensive survey of the quantitative spatial literature, we refer the interested reader to Redding and Rossi-Hansberg (2017), and to Nagy (2022) who focuses on the use of quantitative spatial models to address historical questions.

Section 3 establishes a series of empirical facts which motivate the structure of the model (Section 4). Finally, Section 5 takes the model to the data to disentangle the mechanisms behind the long-run dynamics of cities, and Section 6 concludes.

1 Context

This section provides a brief overview of Britain's historical development over the past three centuries, focusing on aggregate and more local factors that led to the sustained increase in economic growth in the short- and also the long-run.

Figure 1. The industrial revolution in Britain.



(a) GDP

(b) Employment share

(c) Trade openness

(d) Urbanization

Notes: in Panel (a), growth figures are from Broadberry et al. (2015). In Panel (b), employment shares are classified according to the PST system described in Wrigley (2010). We report the available data for male adults in England and Wales (Shaw-Taylor and Wrigley, 2014). In Panel (c), openness is defined as the sum of imports and exports as a share of GDP, using Hills et al. (2010) and Broadberry et al. (2015). In Panel (d), urbanization is defined as the share of total population in cities over 5,000 (Bairoch and Goertz, 1986).

The industrial revolution in Britain The industrial revolution can be broadly characterized by four stylized facts: (a) the emergence of new technologies in key sectors leading to sustained increases in growth rates of income per capita; (b) a declining share of employment in agriculture; (c) the growth in domestic and international trade; and, (d) an increasing share of the population living in cities. How each of these fit together, and which are the causal elements in the context of the industrial revolution that first emerged in England, is not completely settled (see, for example, the survey in Clark, 2014).

Figure 1 depicts trends of these four dimensions. While many of the key industrial technologies emerged in the mid-eighteenth century, growth in per capita output accelerated only in the early nineteenth century (panel a). A further puzzle is presented by the share of employment in the secondary sector, which is high as early as 1710 and grows only marginally until the mid-nineteenth century (panel b). Most striking are the dramatic changes in openness, which accelerates after 1820 (panel c), and urbanization, which grows by nearly fifty percentage points over the century (panel d). Over the course of the nineteenth century, Britain grows from a rural economy to urbanization levels that are quite similar to contemporary ones.

Two highly influential hypotheses on the causes of the industrial revolution are formulated in Mokyr (2009) and Allen (2009). For Mokyr, the industrial revolution was driven by the emergence of "attitudes and aptitudes" (Mokyr, 2021)—a respect for entrepreneurs and inventors, and the growth of useful human capital—that begat the enrichment of society. Kelly et al. (2020) further makes the case that the distribution of useful human capital—mechanical workers—across counties in England was key. Hanlon (2021) corroborates this view with a specific focus on the professionalization of invention through the emergence of engineers. Allen, in contrast, emphasizes the demand for new technologies—high wages and low energy costs induced the capital-biased (and labor-saving) technical change that drove the growth in the export-oriented industries (Allen, 2021). Those high wages arose as a result of globalization and the increasing external demand for manufactured goods in which England had a comparative advantage.

The growing external demand for manufactured output caused a shift in modes of production away from partly rural, low-scale, domestic-oriented and water-powered production to urban, specialized, export-oriented and large-scale factories in which steam power dominated.[4] Stokey (2001) quantifies the potential role of trade, energy cost and technical change, and finds that trade explains all the decline in agricultural production,

[4] As Crafts (1989) made clear, the majority of industrial employment in the early nineteenth century industry was small-scale production for local markets. Growth in export-oriented industry was key to the eventual rise in the average standard of living.

over a quarter of the increase in manufacturing, and half of the increase in real wages (a finding qualitatively supported by Harley and Crafts, 2000; Clark et al., 2014).[5]

As a summary, the early nineteenth century sees a huge transformation of the British economy, with: a sharp increase in GDP per capita, a sharp increase in openness to trade, (some) structural change, and mass urbanization. We shed additional light on the patterns of urbanization in the next section.

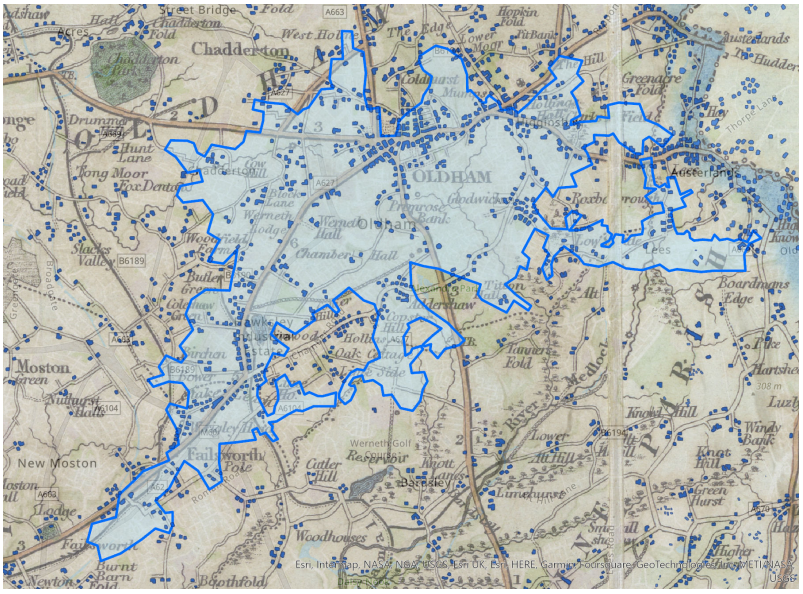
Protoindustrialization, the urbanization of Britain and the rise of cities The shift of industrial production from a rural to an urban setting has been studied extensively, most notably with the hypothesis that rural "protoindustrialization" constituted an important stage in the progress to more general industrialization (Mendels, 1972). That such protoindustrialization caused the industrial revolution has not sustained later analysis (Ogilvie, 2008), but the existence of extensive proto-industries suggests a way to explain the high share of employment in the secondary sector seen in Figure 1 (panel b) in spite of low urbanization levels. In between the low-scale artisanal home-production of finished manufactures and the large-scale, city-centered factory production that typifies late-industrialization are the extensive and well-organized rural industries, some of which exported beyond the locality (Hudson, 2004; Goose, 2014). Early factory production was also frequently rural, relying on water power and with rural workers housed by entrepreneurs around production facilities (Trinder, 2000). While steam engines began to proliferate in the eighteenth century (Nuvolari et al., 2011), the transition to steam engines as the predominant motive power, and the "triumph of the factory system" that went along with it, was not complete until the mid-nineteenth century (Musson, 1976).

One illustration of the swift concentration of manufacturing production over the nineteenth century is Oldham, in the North-East of Manchester. At the onset of the nineteenth century, Oldham does not exist yet as a town: there are numerous hamlets where cottage industry takes place, including one called Oldham, and only a few textile mills appear between 1790–1820. By the end of the nineteenth century, Oldham is a factory town of about 140,000 inhabitants which produces a very significant share of the whole country's textile output (almost a quarter of cotton production in 1911). Figure 2 provides an illustration of such rapid urbanization.

The example of Oldham is emblematic of a widespread, but local, phenomenon whereby smaller settlements would consolidate into towns, small towns would grow fast into

[5]There are other explanations for the change in demand for manufactured output. For instance, Voigtländer and Voth (2013) considers a model in which urbanization and manufacturing demand result from non-homothetic preferences. In the presence of a non-monotonic relationship between income and death rates, the demographic shock of the Black Death causes the economy to shift to a high income (urbanized and industrialized) steady state which prefigured the industrial revolution in Europe.

Figure 2. The urbanization of Oldham.



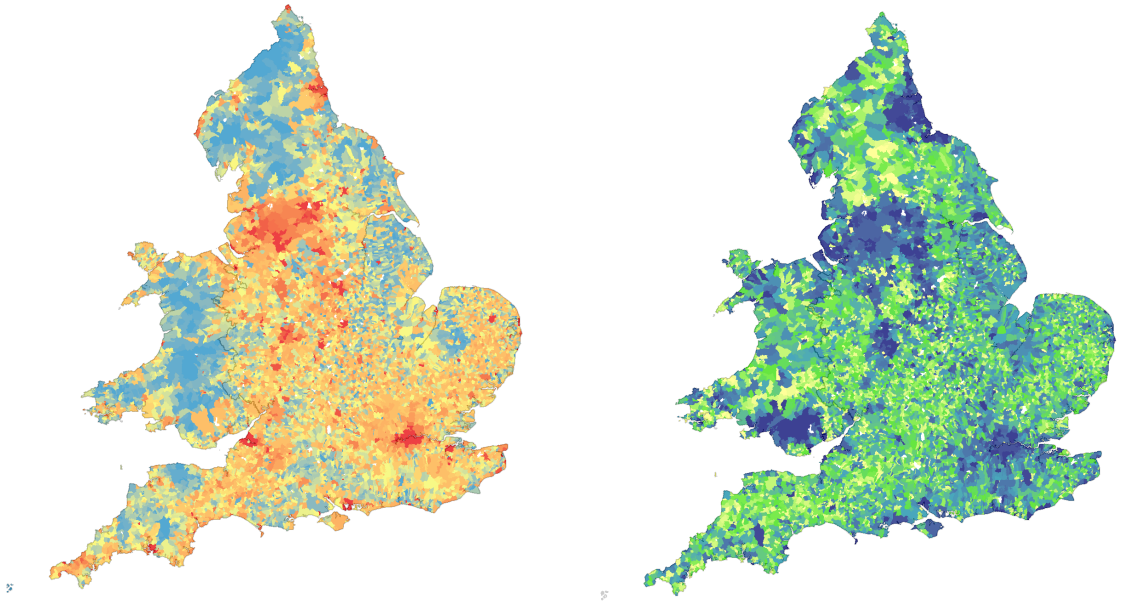
Notes: The underlying map is a county map of Lancashire around 1790–1810 where built-up is indicated by darker blue rectangles. The lighter blue area shows the urban boundaries of Oldham around 1880–1900, as defined by areas of contiguous built-up.

cities with a high concentration of manufactured industries. This pattern is visible within each county of England and Wales. There is, however, a more aggregate geography of employment and growth that is worth discussing. As shown in panel (a) of Figure 3, employment (or population) is highly concentrated in a few regions in the early nineteenth century: Lancashire, North Yorkshire, the West Midlands, Northumberland, London and a few isolated cities (e.g., Bristol, Plymouth). The geography of employment growth closely follows that of initial employment, with a few exceptions (panel (b) of Figure 3). While Lancashire, North Yorkshire, the West Midlands, Northumberland and London grow substantially into large, densely populated regions, the South of Wales also grow significantly (due to a location advantage and proximity to coal).[6] In other regions, growth is more unevenly distributed with a few cities concentrating most of the (new) workforce.

Settlements of the early nineteenth century give way to a dense network of cities by the end of the nineteenth century. We shed some light on the nature of these growth patterns in Figure 4. In panel (a), we show the distribution of employment across cities in 1817 and in 1881 and we see that there is a massive population increase: cities grow by a factor of 3 on average. In panel (b), we show the distribution of specialization, as captured by a 2-digit industry-based Herfindahl index of employment. Cities are mechanically

[6] Shaw-Taylor and Wrigley (2014) suggested that the 19th century shift of manufacturing production into towns outside of London was key to the success in the 19th century.

Figure 3. The geography of employment and growth in Britain.



(a) Employment density (1817)

(b) Growth (1817–1881)

Notes: These two maps represent deciles of population density in 1817 (panel a, as computed from a quasi-census of baptism records conducted between 1813–1820, low: blue, high: red) and deciles of population growth between 1817–1881 (panel b, where population in 1881 is calculated using micro-census records, low: yellow, high: blue). The geographic unit is a parish (see Section 2).

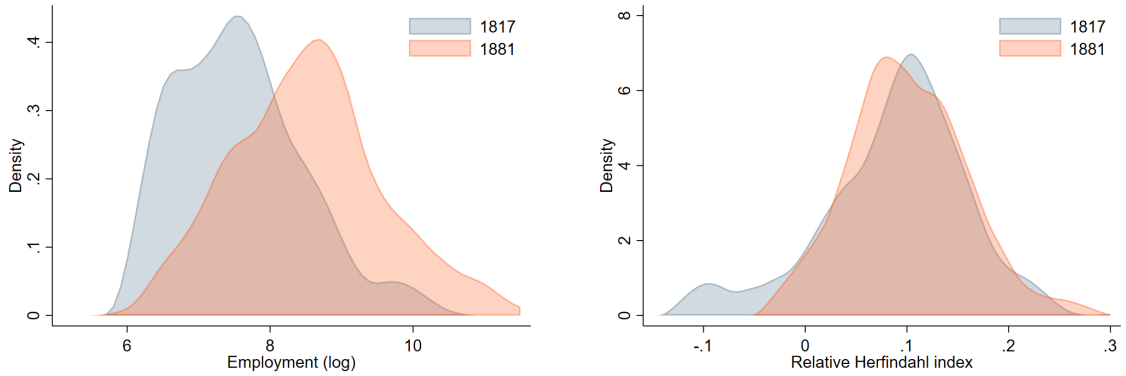
more specialized than the overall economy, but there exists large variation in the degree to which they specialize. We will see later that this variation is tied to the portfolio of location advantages that they hold across the different product varieties. In particular, the trade shock induced a specialization in products where cities hold a comparative advantage, especially so in cities that are most exposed to (internal and external) trade.

Land supply and the local geography of urbanization So far, we have discussed the determinants of urbanization and growth without discussing possible geographic constraints. Industrialization led to rising demand for space within, then around, cities: "Land inside the older towns was acquiring a scarcity value [...] Open spaces inside the older towns vanished rapidly..." (Hoskins, 1988, p.185); "Manufactures ran up their mills, factories and works on the edge of existing towns[...]" (Hoskins, 1988, p.183).

While factory production using new technologies concentrated in growing cities (see, e.g., Trew, 2014) and met the accelerating external demand after the 1820s, the interaction between the growing demand for scale that resulted from trade, and the spatial limits on industrial expansion was key (as Hennessy, 2006, p.103, suggests),

> When size brought economies of scale with the growth of world markets ...

Figure 4. Specialization and growth in cities.



(a) Employment (b) Specialization

Notes: These two figures represent the distribution of employment (panel a) and specialization (panel b) across potential cities of England and Wales (see Section 2 for a definition of these potential cities and their boundaries). The distributions are shown in 1817 (blue) and in 1881 (orange). Specialization is captured by a Herfindahl index calculated as the difference between the city-specific Herfindahl index and a national one:

$$\tilde{h}_c = \sum_i s_{ic}^2 - \sum_i s_i^2$$

where s_{ic} is the employment share in city c and industry i and s_i is the nationwide employment share in industry i. Industries are captured at the 2-digit level.

> a steelworks high up in a south Wales valley or a shipyard crammed into the narrow banks of the lower Tyne experienced increasing disadvantage. Take the great plant at Dowlais ... Tucked between the hills above Merthyr Tydfil almost into the uplands of the Brecon Beacons, its site was hopeless...

Such constraints could be geographic or could be the remnants of historical property rights over land that persisted (Denman, 1958; Hoskins, 1988; Neeson, 1996; Hudson, 2004). In particular, land ownership patterns at the city's fringe shaped the potential for urban expansion. In discussing the constraints on attaining scale in the smelting of iron in South Wales, Trinder (2000) notes that "[p]atterns of housing were dispersed, following patterns set by pre-existing fields and property boundaries rather than those of order and convenience" (p.820).

2 Data

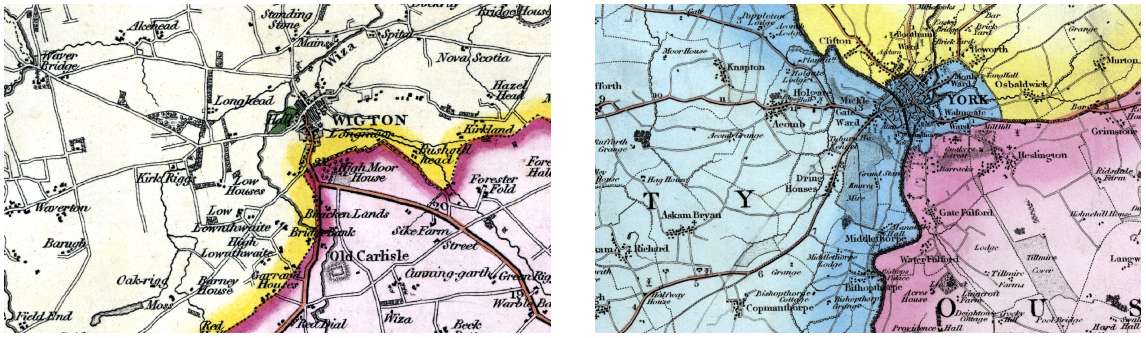
This paper combines data on the evolving geography of England and Wales, the boundaries of cities and the characteristics of their fringe, the transportation network, with about 200 years of data on city composition. In this section, we present our data sources and describe data construction, most notably, the identification of early settlements and of land fragmentation within their fringe.

2.1 Data sources

Historical county maps of England and Wales In the early nineteenth century, the population of Britain mostly lives in rural settlements or small towns. This allows for a significant overhaul of the urban network during the century, with the rise of new cities sprawling around existing hamlets. This also represents a challenge: how can we identify consistent boundaries of potential cities from these early settlements?

Our approach exploits early county maps, drawn between 1790–1820 and covering the whole territory of England and Wales, which we digitize, geo-reference and process through computer vision. We provide an example of these county maps in Figure 5 centered around Wigton (Cumberland) and York (Yorkshire). Identifying objects of interest on these maps is challenging (see, e.g., Combes et al., 2022, for a discussion of map digitization through machine learning). In what follows, we briefly describe our approach to identifying (a) built-up, (b) forests and uncultivated lands, (c) textual information and (d) roads of different size/use.

Figure 5. A rich set of historical county maps (1790–1820) displaying built-up.



(a) Wigton (Cumberland) (b) York (Yorkshire)

Note: This Figure displays the surroundings of Wigton in Cumberland (panel a) and York in Yorkshire (panel b).

To identify land use and textual information, we use a standard image segmentation algorithm, called "Quickshift", which groups pixels by their proximity in actual space (i.e., along the physical distance) and in the space as defined by colors. The procedure is disciplined by three parameters: a scale parameter (calibrated on the size of the minimal object of interest, e.g., one farm), a maximum physical distance which disciplines the extent to which the algorithm looks for neighbors, and a relative weight between distance in the multi-bands-space and physical distance. This first step produces a set of superpixels with different shapes and colors. The second step consists in characterizing the color, texture and shape of these superpixels with a set of variables: average R/G/B intensities, within-standard deviation in R/G/B intensities (e.g., to capture stripes), the compactness, area, perimeter and complexity of the superpixel. The third step consists

in producing a training sample, i.e., a small set of labeled superpixels (in a few categories: text, forest, sparse trees, built-up, etc.). The fourth step trains a random forest algorithm on the few labeled superpixels and predicts land use for all other non-labeled superpixels (Combes et al., 2022). Our procedure leads to a very precise classification of built-up (our main object of interest), with a precision of about 0.95.

To identify roads, we rely on a less standard approach: thin, non-straight lines are notoriously difficult to detect using computer vision. Our approach will rely on the nature of roads: they are black lines on the map, but, in practice, they are designed to best facilitate transportation across space. In that respect, they markedly differ from all other black lines on the maps, e.g., the text or gradient lines in mountainous terrain.[7] We exploit this discriminatory feature as follows: (i) we create a "stylized travel cost" matrix for each map, penalizing travel outside of darker, black pixels (see Appendix B.1); (ii) we draw many random departure points from black pixels on the image and compute the least cost paths to each other pixels of the image for each draw; (iii) for each draw, we then randomize many arrival points and start drawing the actual minimum paths between departure and arrival pixels. The resulting picture shows the intensity to which each pixel is used along the many, different travel routes. As shown in Appendix B.1, this algorithm does not only identify roads, but also their respective importance within the network.

Census of England and Wales The main data source is the Census of England and Wales, which provides a unique characterization of population and industrial composition at the level of about 11,500 parishes over the course of two centuries (1801–1911, 1971–2011). The census provides population counts from 1801 onward, but a precise decomposition of the labor force across occupations only after 1851 (when the micro-census records become available). We thus rely on a quasi-census based on (adult male) Anglican baptism records collected between 1813 and 1820 (referred to below as 1817) in order to retrieve consistent industrial composition at the parish level before the time of rapid industrialization (Shaw-Taylor and Wrigley, 2014). One issue with census data is that the lowest administrative units—the parishes—are regularly redefined, merged or split over the course of the nineteenth century. We thus apply an "envelope" algorithm which considers the union of the different parishes covering the same points over time (see Appendix B.1).

[7] Rivers and county boundaries may satisfy similar properties as roads. We can however clean these false positives using other sources of data cataloging them separately.

Geography and land ownership To characterize the immediate neighborhood of cities and the local, temporary drivers of urban land supply, we gather high-quality raster maps at a disaggregated level: elevation (OpenLandMap, 30m resolution); soil organic carbon content (OpenLandMap, 250m resolution); soil bulk density (Soil bulk density, 250m resolution); a detailed soil classification (National Soil Resources Institute); and a dataset of all rivers and smaller streams in England and Wales.

A crucial component of the empirical analysis consists in the construction of an exogenous measure of land fragmentation based on topography and soil characteristics (i.e., natural breaks between possible agricultural land parcels). One channel through which exogenous land fragmentation might put a strain on city growth is that it might contribute to fragmented land ownership and thus make the land harder to assemble. To validate this channel, we collect actual measures of land ownership fragmentation at the beginning of the nineteenth century (from land tax registers) and in 1851, 1861 and 1881 from micro-census records where land acreage is reported by landowners. Inferring land ownership fragmentation from micro-census records requires a systematic text analysis as the information has not been coded by the I-CeM project.

Transportation We complement the previous data on early settlements, population, occupation and geography with the transportation infrastructure (main roads, navigable waterways, train lines and train stations), as provided by the Cambridge Group for the History of Population and Social Structure. This dynamic characterization of transportation allows us to measure access to resources through the transportation network and trading costs across different cities (see Appendix B.1).

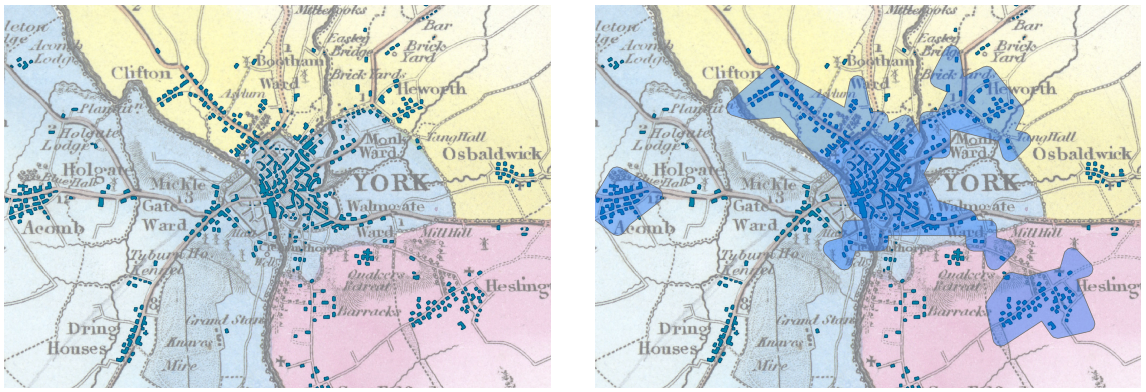
2.2 Data construction

Our empirical strategy exploits two sources of variation to predict late-nineteenth century population and industrial specialization. We first describe how we construct a predictor ζ_c for the population of city c that is exogenous to later city dynamics. The idea is to identify (i) the historical boundaries of an urban settlement, c, before the time of rapid industrialization, and (ii) land fragmentation at its immediate fringe as induced by local gradients in soil conditions.

Urban settlements at the onset of rapid industrialization In Section 2.1, we describe how we identify built-up areas around 1790–1820, at the onset of rapid industrialization. At the time, however, isolated built-up is a frequent occurrence and there are many very small settlements (see, for instance, the surroundings of Oldham in Figure 2).

The settlements that would consolidate into towns and grow into cities are large, dense or spread across contiguous hamlets. To identify future cities and their early boundaries, we develop a procedure to select urban settlements as sufficiently large clusters of built-up areas.

Figure 6. Clustering procedure.



(a) Detecting built-up

(b) Clusters and city boundaries

Note: the blue rectangles in panels a and b are built-up areas detected by our algorithm (see Section B.1); the blue line in panel b is the outcome of our clustering algorithm adapting De Bellefon et al. (2021).

Figure 6 illustrates our approach. We first detect built-up with the previous algorithm (panel a). We then follow the procedure developed in De Bellefon et al. (2021) to identify nucleus of high density and contiguous areas of excess building density. We illustrate the outcome of this procedure for the city of York and its neighboring towns in panel (b) of Figure 6.[8] We then construct predicted boundaries for these cities at the end of the nineteenth century by assuming that towns and cities all grow at the same proportional rate across the country and do so homogeneously in any direction. In practice, towns and cities all expand to some degree, but this expansion is not homogeneous across all directions and not homogenous across cities. We discuss next how we predict the extent of such expansion, and therefore cities' late-nineteenth century population, with land ownership fragmentation.

Land fragmentation around city boundaries Land ownership fragmentation across different parts of England and Wales was not only instrumental to the development of agriculture, as illustrated by the effect of enclosures (Neeson, 1996), but it was also crucial in disciplining city growth during the era of rapid industrialization.

Indeed, when land markets are not perfectly competitive and land parcels and their rights of use cannot be split arbitrarily, developing land at the fringe of cities may be a

[8] An alternative procedure is described in Arribas-Bel et al. (2021).

challenge. For instance, a textile mill requires a large parcel of flat land to construct a factory, but also possible access to water sources. When a suitable location spans multiple land parcels, the possible buyer needs to engage in multilateral bargaining in which the value of the marginal parcel increases as the buyer acquires rights to use for other parcels. The number of different parties then matters. This issue is a "standard" hold-up problem, which has been labeled as the land assembly problem in this specific context (see, e.g., Eckart, 1985; Strange, 1995). The consequences for city growth are straightforward: high land fragmentation at the fringe of the city makes negotiations to develop the land for urban use costly. As a result, cities with fragmented ownership in their immediate fringe have a less elastic land supply in the short run and may be worse at responding to sudden bursts in land demand. The rapid industrialization induced by technological progress and a surge in external trade in the mid-19th century, as evidenced in Section 1, is one such shock.

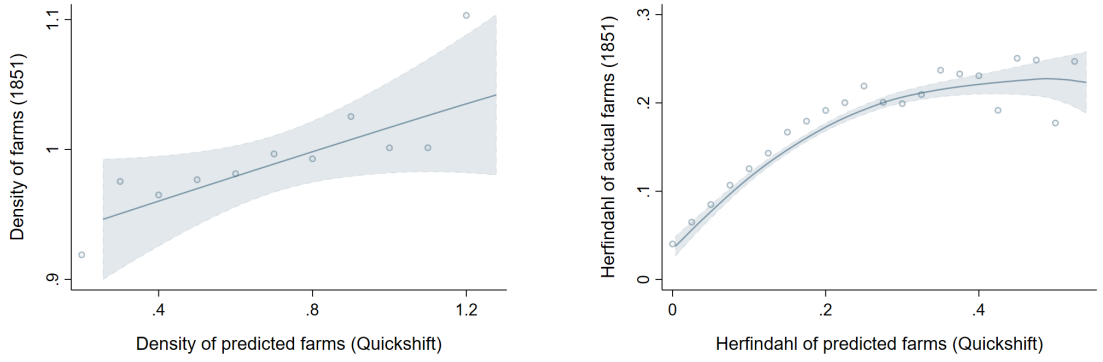
A major issue with land ownership fragmentation in the immediate fringe of cities is that it may reflect past and future city dynamics. The relationship between land ownership fragmentation and city growth may thus be "contaminated" by omitted variation—the relative productivity of land between urban and rural use inducing different structure of ownership—but also reverse causality—land owners less willing to own large plots around cities most likely to expand. For these reasons, we would like to consider a measure of land ownership fragmentation with the following characteristics: (i) the measure, as evaluated within a neighborhood of city boundaries at the beginning of the nineteenth century, should predict city growth in the following decades; (ii) the measure should not directly affect the later evolution of cities during the twentieth century, when conditioned on the right control variables (e.g., the elasticity of land supply in later periods, once cities have expanded beyond the narrow ring and are thus subject to another topography at their borders).

We construct a plausibly exogenous measure of land fragmentation by exploiting fine-grained terrain characteristics including elevation, ruggedness, time-invariant soil attributes and water bodies. We leave the details of the procedure to Appendix B.2 and only summarize its main steps below. First, we combine these different dimensions of characteristics into a multi-band raster covering England and Wales at a resolution of 30m (see the left panels of Appendix Figure B5 around Darlington and Barnard Castle). Topography and soil display local breaks which we illustrate in Appendix Figure B5 with yellow lines. Second, we use, again, our unsupervised segmentation algorithm which isolates homogeneous (and contiguous) color zones. In our previous application, the multi-band raster was constituted of three bands (R: red, G: green, B: blue). In the present one, the raster may contain many more bands, but the principle is the same: the

algorithm maximizes a weighted sum of target distances within constituted superpixels, with a weight allocated to physical distance relative to the "value distance." A superpixel with similar values is, here, a patch of land with homogeneous topography and soil characteristics: a typical agricultural parcel, e.g., as delineated by enclosures in some parts of England and Wales.[9] Third, we compute the density of predicted farms in the fringe of cities by drawing buffers of different widths (e.g., 1, 2, 3, 5 kms) around city boundaries at the onset of rapid industrialization. One can think about the narrow rings as predicting the propensity for cities to grow over the nineteenth century and the wider rings as controlling for later land supply elasticities. The quantitative model developed in Section 4 will allow for cities to face varying land supply elasticities over time, in part to capture the previous intuition.

Validating our measure of land fragmentation The predicted measure of land fragmentation should be correlated with the actual fragmentation of land ownership. We validate the predicted measure of land fragmentation by comparing it with actual farm density as collected from micro-census records in 1851 across all parishes of England and Wales (see Figure 7, panel a). We also provide a test based on the actual and predicted concentrations of farm ownership (panel b).

Figure 7. Validation of the predicted measure of land fragmentation.



(a) Density (b) Land concentration

Notes: The left panel displays the measure of predicted fragmentation versus actual farm density as collected from micro-census records in 1851 across all parishes of England and Wales. We create 20 bins of density and the dots represent the average actual farm density within each bin. The lines are locally weighted regressions on all observations. Note that the conditional correlation between the two measures is 0.25 (once conditioned on the separate topographic and soil characteristics). The right panel repeats the same exercise with the actual and predicted concentration of farm ownership.

[9]Within the class of local mode-seeking algorithms able to perform this classification, we opt for Quickshift and we calibrate: the scale parameter and the maximum physical distance to capture the size of a typical farm; and the relative weight between distance in the multi-bands-space and physical distance to discipline how compact the predicted farms will be.

While this exercise shows that our measure of land fragmentation does predict land ownership fragmentation, it does not rule out that the measure is related to city dynamics before industrialization. Thus, we further validate the land fragmentation measure by comparing urban settlements with different degrees of land fragmentation in their "external crusts," as calculated at the onset of the nineteenth century (1817). Appendix Figure B8 provides the equivalent of a balance test comparing population density and industrial specialization (calculated as a Herfindahl index across 3-digit occupations) for settlements with above- and below-median predicted land fragmentation in their immediate fringe. We do not find marked differences, suggesting that settlements with different degrees of land fragmentation were similar in their population density and industrial specialization before the era of rapid industrialization.

Shift-share predictions of industrialization and specialization To predict industrial specialization in city c, we rely on city-specific employment shares across 2-digit sectors $i \in \{1, \dots, I\}$ in 1817, s_{ic}, and aggregate employment growth across sectors, $g_i > 0$. We first construct a predictor for urban employment growth, $\hat{g}_c$,

$$\hat{g}_c = \sum_i s_{ic} g_i$$

based on aggregate industry-specific shifts and the initial composition of employment within the predicted urban area. We then construct a predictor for industrial specialization $\hat{h}_c$, as a Herfindahl index using predicted employment shares based on aggregate industry trends:

$$\hat{h}_c = \frac{\sum_i (s_{ic} g_i)^2}{\left(\sum_i s_{ic} g_i\right)^2}$$

These shift-share predictions constitute one building block of our empirical approach. We describe the empirical strategy next.

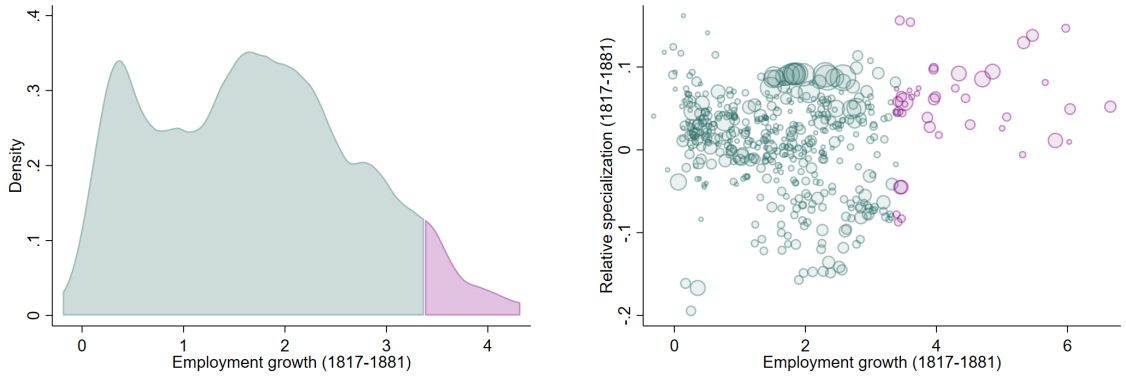
3 Empirical facts

In this section, we document the following empirical facts: 1. By the end of the nineteenth century, there are many cities, and they are very heterogeneous in their industrial mix; 2. Land fragmentation around initial city boundaries and our shift-share predictor of specialization, respectively, predict population and industrial specialization at the end of the nineteenth century; 3. Specialization leads to poor city performance in the long run.

3.1 The rise of (different) cities

We first provide a few descriptive statistics illustrating the heterogeneous rise of Great British cities, experiencing very different population growth and industrial specialization.

Figure 8. Fast-growing cities and slow-growing cities grow more or less specialized.



(a) Employment growth (b) Specialization

Notes: The left panel displays the distribution of employment growth within the (larger) boundaries of about 500 towns. The top decile of growing settlements are highlighted in purple. In the right panel, we show the relationship between employment growth and the evolution of industrial specialization across urban areas.

The heterogeneous rise of Great British cities By the end of the nineteenth century, urbanization is slowing down in Britain and some cities are specialized in a few industries while others were more diverse, a process disciplined by geography, trade and the distribution of their location advantages. We now shed some light on the heterogeneous rise of Great British cities.

Our procedure to detect urban settlements around 1790–1820 (see Section 2.2) identifies more than 500 potential cities, their initial boundaries, and their predicted boundaries by the end of the nineteenth century. One issue is that urban settlements change in their area during the period and they do so endogenously. To strike a balance between comparing employment numbers between similar areas across time and dealing with endogenous urban sprawl, we associate employment within each 2-digit industry in 1817 and in 1851–1911 (and in 1971–2011) to a city c by intersecting the predicted boundaries around 1880–1900 with parish boundaries, allocating employment using the area share of the intersection. In summary, we construct all our employment/population variables within expanded cities (i.e., with a buffer around the delineation presented in Figure 6).[10]

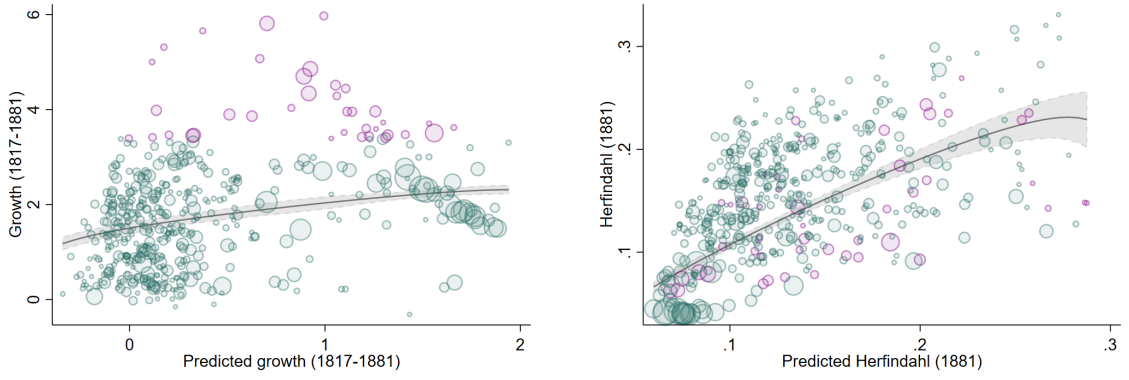
[10]Throughout the empirical analysis, we use "city employment" and "city population" interchangeably.

Figure 8 shows the distribution of employment growth across our urban settlements. We see that the median settlement grows by a factor of about 3. The growth rate for the top decile (in purple) is above 300% (panel a). The large heterogeneity in employment growth is also reflected in a large heterogeneity in industrial specialization. The fastest-growing cities become very specialized (panel b), an effect which is entirely explained by the fact that fast-growing cities are initially specialized across a few key sectors that flourish over the course of the nineteenth century (see Appendix B.3).

The overall role of geography, trade and initial conditions By the end of the Industrial Revolution, cities of England and Wales have widely different economic structures. A key factor of specialization is the initial industrial mix interacted with the gradual opening of the economy to (internal and external) trade: some cities have initial conditions that lead to more specialization, others remain structurally more diverse. In what follows, we further restrict the sample to the top-50% larger cities of our sample in 1817—this leaves us with about 285 cities.

We illustrate the role of initial industries, as proxies for initial location advantages, in Figure 9 where we compare the actual employment growth, g_c, and industrial specialization, h_c, in 1881 with their predictions, $\hat{g}_c$ and $\hat{h}_c$, as computed using a "shift-share" design based on initial industries in 1817.

Figure 9. The role of the initial industrial mix.



(a) Employment growth (1881)

(b) Specialization (1881)

Notes: The left panel displays the measure of predicted employment growth, $\hat{g}_c$, against the actual employment growth. The top decile of growing settlements are highlighted in purple. The line is a locally weighted regression on all observations. The right panel repeats the same exercise with the actual and predicted industrial concentration.

Figure 9 shows that the initial industrial mix strongly predicts employment growth within urban areas but also, and mostly, industrial specialization. While the measure of predicted employment growth, $\hat{g}_c$, explains 8% of the variation in actual growth, pre-

dicted specialization, $\hat{h}_c$, explains 25% of the variation in actual specialization.[11]

To better illustrate the respective role of initial conditions and geography in explaining urban dynamics across urban settlements, we run simple regressions with employment growth and specialization by 1881, adding sequentially controls for initial conditions (employment levels and specialization), shift-share predictors (constructed from the initial industrial mix and based on 2-digit industries), and a set of geographic indicators.

Table 1. The role of geography, trade and initial conditions.

Adjusted R-squared	Employment growth (1881)	Specialization (1881)
Initial conditions (1817)	0.234	0.319
+ Prediction (industry)	0.336	0.470
+ Geography	0.609	0.756
Observations	285	285

Notes: Initial conditions are initial employment and Herfindahl in 1817; the predictions are industry-based shift-shares; and geography includes: slope, elevation, bulk density, carbon content, latitude/longitude, travel time to coal, travel time to London, distance to/density of roads (1830), distance to/density of stations (1846), distance to waterways distance to the shore, share of arable, suitability to grow wheat and grass.

Table 1 shows that initial conditions, as captured by initial employment and the Herfindahl index in 1817 and by the industry-based shift-shares, are important in explaining the heterogeneous rise of Great British cities. However, geography is also instrumental and explains a similar share of the overall variation. There are three types of geographic factors which predict urban development: (i) market access and connectivity; (ii) local agricultural conditions (see, e.g., Coeurdacier et al., 2022; Heblich et al., 2020); and (iii) local topography and other factors which could influence land supply at the fringe of existing urban settlements. In particular, some cities are constrained in their growth, others are not. Next, we describe the predictive power of our instruments, including a measure of land fragmentation to capture the latter effect.

[11]Both our actual and predicted indices of specialization are based on 2-digit industries. We provide insights about the respective role of specialization across 1-digit industries (such as manufacturing and services) versus specialization within 1-digit industries in Appendix B.3. The latter explains about 60% of the relationship between actual and predicted specialization. In unreported robustness checks, we verify that our main results of Section 3.3 are robust to controlling for a Herfindahl index computed at the 1-digit level.

3.2 Exogenous variation in population and specialization (1881)

To establish a causal link between cities' population and specialization and their subsequent dynamics, we need to isolate exogenous variation among the numerous factors shaping the transformation of British cities during the nineteenth century (geography, trade and initial conditions). We do so by focusing on two dimensions characterizing nineteenth-century urban settlements which (i) were instrumental in explaining the swift rise of cities during the mid-nineteenth century, and (ii) are unlikely to act as significant constraints for later economic development.

To isolate exogenous variation in cities' late-nineteenth century *population*, we use our exogenous proxy for land ownership fragmentation in the immediate fringe of urban settlements, ζ_c.[12] As already discussed in Section 2.2, urban settlements were growing rapidly during the era of mid-19th century industrialization, leading to a massive increase in demand for urban land. A high degree of ownership fragmentation could slow down negotiations to acquire additional land for urban use.[13] To isolate exogenous variation in cities' late-nineteenth century *specialization*, we exploit our "shift-share" predictor for specialization ($\hat{h}_c$), while always controlling for the predictor for growth ($\hat{g}_c$).

We explain urban development in city c, $\mathbf{y}_c^{81}$, with these two exogenous predictors, $\mathbf{p}_c$, as well as a large set of controls, $\mathbf{X}_c$,

$$\mathbf{y}_c^{81} = \alpha + \beta \mathbf{p}_c + \gamma \mathbf{X}_c + \varepsilon_c$$

and we report the main estimates of interest in Table 2. First, and consistent with Figure 9, predicted specialization does induce more industrial specialization in 1881. The passthrough between the predicted Herfindahl index ($\hat{h}_c$) and the actual one (h_c) is around 0.26. Second, we find that predicted land fragmentation impacts the capacity of cities to grow during the era of industrialization, and hence their population in the late nineteenth century. We find that one standard deviation in land fragmentation (about 0.32), ζ_c, decreases 1881 city population by about 20% (0.32×0.63)—an effect that is far from negligible.

[12] The role of agricultural hinterlands in fueling urban growth has been discussed in Matsuyama (1992), and more recently in Coeurdacier et al. (2022), Heblich et al. (2020) and Nagy (2023). Our instrument relies on another theoretical mechanism, the land assembly problem, and we show in a robustness check that we can control for local agricultural productivity (Turner, 1982) and the modernization of agriculture in hinterlands (Caprettini and Voth, 2020).

[13] A recent contribution discusses the role of enclosure acts on crop yields and land inequality (Heldring et al., 2022). Our source of variation to explain land fragmentation is, in essence, orthogonal to the one used in their analysis. However, we do control for the specific role of enclosures in a robustness check.

Table 2. Predicting cities' late-nineteenth century population and specialization.

	Herfindahl (1881)	Population (1881)
	(1)	(2)
Predicted Herfindahl ($\hat{h}_c$)	0.262	-3.456
	(0.052)	(1.780)
Land fragmentation (ζ_c)	0.006	-0.630
	(0.005)	(0.152)
Observations	285	285

Notes: The set of baseline controls include initial employment and Herfindahl in 1817, and: slope, elevation, bulk density, carbon content, latitude/longitude, travel time to coal, travel time to London, distance to/density of roads (1830), distance to/density of stations (1846), distance to waterways, distance to the shore, share of arable, suitability to grow wheat and grass, and the "shift-share" predictor of urban employment growth ($\hat{g}_c$) defined in Section 2.2. Population (1881) is cities' (log) employment in 1881. Herfindahl index (1881) is based on 2-digit industries (sample mean: .14). The predicted Herfindahl index in 1881 is the industry-based "shift-share" defined in Section 2.2 (sample mean: .15).

3.3 Specialization, long-run dynamics and the death of (some) cities

We now use the previous exogenous variation to estimate how cities' population and specialization influence their long-run dynamics. We regress the share of unskilled workers in 1971 in city c, s_c^{71}, on urban development in 1881, $\mathbf{y}_c^{81}$, conditioning on a large set of controls, $\mathbf{X}_c$,

$$s_c^{71} = \alpha + \beta \mathbf{y}_c^{81} + \gamma \mathbf{X}_c + \varepsilon_c$$

and instrumenting $\mathbf{y}_c^{81}$ with our exogenous predictors $\mathbf{p}_c$.[14]

Table 3 reports the results, and confirms that industrial specialization is detrimental in the long run. In column (1), we control for a wide set of confounders and find that cities with a one standard deviation higher value of the Herfindahl index in 1881 have a 1.415 standard deviations higher share of unskilled workers in 1971. Larger cities in 1881 also tend to have slightly worse economic outcomes in the long run. However, the coefficient on population is only marginally significant, and the magnitude of the standardized coefficient is only about 60% of the effect of specialization (0.918). Overall, we conclude that industrial specialization is a key driver of British cities' long run performance.

The effect of specialization on long-run city performance might reflect two distinct forces. First, nationwide industry decline might hurt cities specialized in the declining industries. Second, specialization might have a direct effect on future city growth, à la Jacobs (1969). In column (2), we aim at cleaning our estimates from the former effect

[14]We consider the share of unskilled workers in 1971 (derived following Heblich et al., 2021), before the swift decrease in manufacturing employment accompanying Thatcher's reforms.

Table 3. The long-run effect of city population and specialization.

	Unskilled workers (1971)		
	(1)	(2)	(3)
Herfindahl (1881)	0.762	0.779	0.580
	(0.277)	(0.275)	(0.342)
	[1.415]	[1.447]	[1.078]
Population (1881)	0.018	0.017	0.018
	(0.009)	(0.011)	(0.013)
	[0.918]	[0.850]	[0.913]
Observations	285	285	278
Growth shift-share (1817–1881)	Yes	Yes	Yes
Growth shift-share (1881–1971)	No	Yes	Yes
County fixed-effects	No	No	Yes

Notes: The set of baseline controls include initial employment and Herfindahl in 1817, and: slope, elevation, bulk density, carbon content, latitude/longitude, travel time to coal, travel time to London, distance to/density of roads (1830), distance to/density of stations (1846), distance to waterways, distance to the shore, share of arable, suitability to grow wheat and grass, and the "shift-share" predictor of urban employment growth ($\hat{g}_c$) defined in Section 2.2. Population (1881) is cities' (log) employment in 1881. Herfindahl index (1881) is based on 2-digit industries (sample mean: .14). The predicted Herfindahl index in 1881 is an industry-based "shift-share" (sample mean: .15). Unskilled workers is the share of unskilled workers in 1971 (sample mean: .88). Standardized coefficients are reported in brackets.

by controlling for aggregate industry trends between 1881–1971 in a shift-share design. Our findings are robust to such industry trends. They are also robust to controlling for county fixed-effects (column 3), thereby exploiting the heterogeneous rise of cities within each of the more than 50 counties covering England and Wales.

Robustness checks We provide a series of robustness checks to support our main empirical findings. First, our identification narrative is that land fragmentation affects urban development only through land supply at the fringe of early urban settlements; this narrative rules out the existence of competing mechanisms arising, e.g., from agricultural productivity (Coeurdacier et al., 2022) or agricultural mechanization at the fringe of cities (see Caprettini and Voth, 2020, with the use of threshing machines) or from the emergence of different informal institutions in places with different structure of land ownership. To alleviate concerns about the first competing channel, we control for the share of arable in local production (which defines the exposure to the repeal of the Corn Laws; see Heblich et al., 2020) and for the potential yield of the most common crops in our baseline specification. We further control for agricultural productivity (using data from Turner, 1982) and agricultural mechanization in parishes surrounding the city. To alleviate concerns about the second competing channel, we control for proxies of local

institutional change (Heldring et al., 2022) and provide detailed balance tests in 1817 across cities with different land fragmentation in their fringe, comparing their density and industrial structure (Section B.3). Section B.5 presents the results of these checks and confirms that they all leave our headline empirical finding, the detrimental long-run effect of specialization, unchanged.

Second, one might wonder whether the results are driven by any of the large nineteenth-century industries. If that were the case, that would affect the interpretation of our results, suggesting that it is primarily specialization in that particular industry, rather than overall industrial specialization, that hurts long-run growth. To address this question, Section B.5 re-runs the empirical specification of column (2) in Table 3 after dropping each of the major industries, one by one. We find that the long-run effect of industrial specialization remains positive in all of these specifications, suggesting that none of the major industries are driving the estimates.

Third, the validity of our instrument requires that it has no direct effect on urban development after 1881. In unreported checks, we show that the vast majority of cities grow out of this narrow ring by the end of the nineteenth century, that we can control for the marginal land supply from the end of the nineteenth century onward, and that there are no differential zoning policies at the fringe of cities—depending on their initial land fragmentation (e.g., green belts, or social housing policies). We also control for access to resources in our baseline specification to reduce concerns that energy is the main factor explaining the rise and then fall of Great British cities.

Finally, we conduct a large number of unreported robustness checks around the baseline specification(s): with different buffers around the city; with different inference to account for spatial correlation; and with different cut-offs to define urban settlements.

The stylized facts presented in this section provide evidence about the joint dynamics of urbanization and specialization in cities, showing that the fate of cities is tightly related to that of their industries. The next section provides a more structural approach, by developing a quantitative model of cities and their industries over time which captures both the spatial linkages across cities and their industries in a given period, as well as possible intertemporal externalities.

4 A multi-sector dynamic spatial model

To provide a framework in which we can study the determinants and long-run consequences of local specialization and development, we develop a spatial, multi-sector, dynamic model of cities. We offer an algorithm to solve the model efficiently and simulate the model on a stylized geography to illustrate how it can replicate the empirical

facts documented in Section 3.

4.1 Setup

The model involves a finite number of cities $c \in \{1, \dots, C\}$ and industries $i \in \{1, \dots, I\}$. Time is discrete and is indexed by t. Within each industry, every city produces its own variety that consumers view as different from the varieties produced in other cities. There is an exogenous number $\bar{L}$ of workers in the economy. Each worker lives for one period, maximizing her utility from the consumption of varieties. Each worker is endowed with one unit of labor. The worker decides which industry to work for and which city to live in.

In what follows, we describe the four main building blocks of the model: workers' preferences, the production technology, the equilibrium within a time period t, and the dynamic process that links subsequent periods to each other.

Preferences If a worker m who lives at time t decides to work in industry i and to live in city c, she chooses her consumption levels to maximize her utility,

$$U_{ict}^m = \max a_{ct}^m \left[\sum_{j=1}^{I} \left(\sum_{d=1}^{C} \left(q_{jdt}^m \right)^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

subject to the budget constraint,

$$\sum_{j=1}^{I} \sum_{d=1}^{C} p_{jdt} q_{jdt}^m \leq w_{ict} + R_{ct}, \tag{2}$$

where a_{ct}^m denotes the level of amenities enjoyed by the worker in city c, while the rest of the worker's utility is drawn from her consumption of industry-varieties. More precisely, q_{jdt}^m denotes the worker's consumption of the city-d variety in industry j, p_{jdt} denotes the price of this variety, w_{ict} is the wage that prevails in the city-industry, and R_{ct} is the worker's share of land rents.[15] When choosing her city, industry and consumption, the worker takes amenities, prices and wages as given. We assume that varieties are substitutes, and they are more substitutable within than across industries, implying $\epsilon > \sigma > 1$.

The worker's amenity level in a given city is a combination of three factors:

$$a_{ct}^m = \bar{a}_c L_{ct}^{-\lambda} \varepsilon_{ct}^m. \tag{3}$$

[15] We assume that rents are redistributed to workers living in the city.

$\bar{a}_c$ is the fundamental amenity level of city c, which may stem from natural characteristics such as climate. $L_{ct}^{-\lambda}$ is a congestion externality that makes cities with a larger population L_{ct} less pleasant places to live. Finally, ε_{ct}^m is an idiosyncratic taste shock for city c that is drawn from a Fréchet distribution,

$$Pr\left[\varepsilon_{ct}^m \leq z\right] = e^{-z^{-1/\eta}}, \tag{4}$$

independently across workers, cities and time periods. If η is high, then workers' utility is influenced to a large extent by their idiosyncratic tastes for cities. As a consequence, they are likely to settle in a city they like, rather than in a city that offers them good economic opportunities. The same would be true if workers faced high costs of moving across cities. Thus, η can also be interpreted as a parameter driving the severity of mobility frictions across cities.

Technology Varieties are produced by perfectly competitive firms. The representative firm producing the city-c variety in industry i at time t faces the production function

$$Y_{ict} = \tilde{\gamma}\tilde{T}_{ict}L_{ict}^{\gamma}H_{ict}^{1-\gamma}, \tag{5}$$

where $\tilde{T}_{ict}$ is the TFP of industry i in city c at time t, L_{ict} is the number of workers hired by the firm, and H_{ict} is the amount of land used by the firm. We define the constant $\tilde{\gamma} = \gamma^{-\gamma}\left(1-\gamma\right)^{-(1-\gamma)}$ in order to simplify the subsequent formulas.

Varieties can be traded across cities, but they are subject to iceberg trade costs. We denote the iceberg trade cost prevailing between cities c and d in industry i at time t by τ_{icdt}. Naturally, we assume that trade costs are always non-negative, which amounts to $\tau_{icdt} \geq 1$.

Land is supplied in each city according to the supply function

$$H_{ct} = r_{ct}^{\zeta_{ct}-1}, \tag{6}$$

such that r_{ct} is the land rent and $\zeta_{ct} - 1$ is the land supply elasticity. We let the exogenous parameter driving this elasticity, ζ_{ct}, vary both across cities and over time—this feature is essential to mirror the heterogeneity across cities uncovered in the empirical analysis. It is natural to assume that $\zeta_{ct} \geq 1$, i.e., the supply function is never downward-sloping. Land rents are fully redistributed to workers who live in the city.

Within-period equilibrium Before we turn to presenting the dynamic evolution of sectoral TFP levels, we set up the equilibrium within a given time period t for *given* TFP

levels in that period. In this within-period equilibrium, we impose that the labor market clears in each city, the land market clears in each city,

$$\sum_{i=1}^{I} \frac{1-\gamma}{\gamma} w_{ict} L_{ict} = r_{ct} H_{ct}, \tag{7}$$

markets clear for each variety,

$$\left(w_{ict} + R_{ct}\right) L_{ict} = \sum_{d=1}^{C} \left(\frac{P_{idt}}{P_{dt}}\right)^{1-\sigma} \left(\frac{p_{ict}\tau_{icdt}}{P_{idt}}\right)^{1-\epsilon} \sum_{j=1}^{I} \left(w_{jdt} + R_{dt}\right) L_{jdt}, \tag{8}$$

where P_{idt} is the CES price index of industry-i varieties in city d,

$$P_{idt} = \left[\sum_{c=1}^{C} p_{ict}^{1-\epsilon} \tau_{icdt}^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}}, \tag{9}$$

P_{dt} is the CES price index of all consumption goods in city d,

$$P_{dt} = \left[\sum_{i=1}^{I} P_{idt}^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{10}$$

and each worker chooses the combination of city and industry that offers them the highest utility.

Dynamic evolution of TFP We now present the assumptions on how sectoral TFP levels evolve over time. We allow the TFP of each industry to be influenced by agglomeration externalities. In particular, externalities in period t can take the form

$$\tilde{T}_{ict} = T_{ict} L_{ct}^{\alpha} f_i \left(L_{c,t-1}, \left\{L_{jc,t-1}\right\}_{j \in I}\right), \tag{11}$$

where T_{ict} is the exogenous fundamental productivity of industry i in city c at time t. That is, agglomeration externalities may not only depend on the current population of city c (as standard in the literature) as well as on its past population (as in Allen and Donaldson, 2020), but also on the city's sectoral composition in the past. This process, which links the productivity of city-industries to the spatial and sectoral distribution of employment in the previous period, is responsible for the dynamics of the model and, crucially, underlies the joint evolution of cities and industries.

Equation (11) is a flexible formulation of externalities that allows for dynamic within-industry (Marshall–Romer) and cross-industry (Jacobs) externalities. It is a generaliza-

tion of the dynamic TFP process in the one-sector model of Allen and Donaldson (2020).

4.2 Solving the model

In this section, we propose an algorithm to solve for the equilibrium of the model. This algorithm relies on reducing the within-period equilibrium conditions to a system of $3IC$ equations (as shown in Appendix C.1),

$$\begin{aligned} x_{ict}^1 &= \sum_{j=1}^{I}\sum_{d=1}^{C} \left(x_{jdt}^2\right)^{\frac{\alpha+\gamma}{\kappa_{dt}}\frac{\epsilon-1}{\sigma-1}} \left(x_{jdt}^3\right)^{-\left(1-\frac{1+\lambda+\eta}{\kappa_{dt}}\right)} K_{icjdt}^1 \\ x_{ict}^2 &= \sum_{j=1}^{I}\sum_{d=1}^{C} \left(x_{jdt}^1\right)^{\frac{\sigma-1}{\epsilon-1}} K_{icjdt}^2 \\ x_{ict}^3 &= \sum_{j=1}^{I}\sum_{d=1}^{C} \left(x_{jdt}^1\right)^{-\frac{\epsilon-\sigma}{\epsilon-1}} \left(x_{jdt}^2\right)^{-\left(1-\frac{\alpha+\gamma}{\kappa_{dt}}\frac{\epsilon-1}{\sigma-1}\right)} \left(x_{jdt}^3\right)^{\frac{1+\lambda+\eta}{\kappa_{dt}}} K_{icjdt}^3 \end{aligned} \tag{12}$$

where

$$\kappa_{dt} = 1+\lambda+\eta+\left[\frac{1-\gamma}{\zeta_{dt}} - \alpha + \left(\gamma+\frac{1-\gamma}{\zeta_{dt}}\right)(\lambda+\eta)\right](\epsilon-1) \tag{13}$$

is a combination of structural parameters, and the $3\times I\times C$ unknowns x_{ict}^1, x_{ict}^2 and x_{ict}^3 can be obtained from equilibrium prices, wages and population levels through the following change in variables:

$$\begin{aligned} x_{ict}^1 &= P_{ict}^{1-\epsilon} \\ x_{ict}^2 &= w_{ict}^{1-\sigma} L_{ct}^{(\lambda+\eta)(\sigma-1)} \\ x_{ict}^3 &= w_{ict}^{1+\left(\gamma+\frac{1-\gamma}{\zeta_{ct}}\right)(\epsilon-1)} L_{ct}^{1+\left(\frac{1-\gamma}{\zeta_{ct}}-\alpha\right)(\epsilon-1)} \end{aligned} \tag{14}$$

and K_{icjdt}^1, K_{icjdt}^2 and K_{icjdt}^3 are the following functions of exogenous variables:

$$K_{icjdt}^1 = \begin{cases} \left(\frac{1-\gamma}{\gamma}\right)^{-\frac{1-\gamma}{\zeta_{dt}}(\epsilon-1)} \hat{T}_{jdt}^{\epsilon-1}\tau_{jdct}^{1-\epsilon} & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$$

$$K_{icjdt}^2 = \begin{cases} \left(\gamma\bar{U}_t\right)^{1-\sigma} \bar{a}_c^{\sigma-1} & \text{if } c=d \\ 0 & \text{otherwise} \end{cases}$$

$$K_{icjdt}^3 = \left(\gamma\bar{U}_t\right)^{1-\sigma} \left(\frac{1-\gamma}{\gamma}\right)^{-\frac{1-\gamma}{\zeta_{dt}}(\epsilon-1)} \hat{T}_{jct}^{\epsilon-1}\bar{a}_d^{\sigma-1}\tau_{jcdt}^{1-\epsilon}$$

such that

$$\hat{T}_{ict} = T_{ict} f_i\left(L_{c,t-1}, \left\{L_{jc,t-1}\right\}_{j\in I}\right) \tag{15}$$

is the part of TFP that is exogenous in period t.

The solution algorithm consists of guessing an initial distribution of x^1_{ic0}, x^2_{ic0} and x^3_{ic0} in period 0. This is followed by inserting the guesses on the right-hand side of system (12), obtaining an updated guess of x^1_{ic0}, x^2_{ic0} and x^3_{ic0}, and iterating on (12) until convergence. With the equilibrium values of x^1_{ic0}, x^2_{ic0} and x^3_{ic0} in hand, one can express period-0 price indices, wages, population and sectoral employment levels by inverting the system (14). Finally, one can obtain period-1 TFP levels from equation (15) and repeat the procedure for the next period.

4.3 An illustration in a linear economy

In this section, we simulate the model on a simple geography to illustrate how it can rationalize: 1. The rise of (different) cities as influenced by trade and local comparative advantages; and 2. The long-run dynamics and the death of (some) cities, as documented in Section 3.

A linear economy We focus on a country with 200 cities arranged on a line. There are two industries in the country. At the beginning of period 0, the 100 cities to the West of the country centroid have a high TFP of 1.05 in industry 1, and a lower TFP of 0.95 in industry 2. The pattern is reversed in the 100 Eastern cities (top left panel of Figure 10).

Using the algorithm proposed in Section 4.2, we simulate this stylized economy both under autarky and under trade. In either scenario, we simulate the economy for two subsequent time periods, period 0 and period 1. Under autarky, we assume that trade costs between cities are infinitely high. Under trade, we assume that the cost of trading between cities c and d takes the following form,

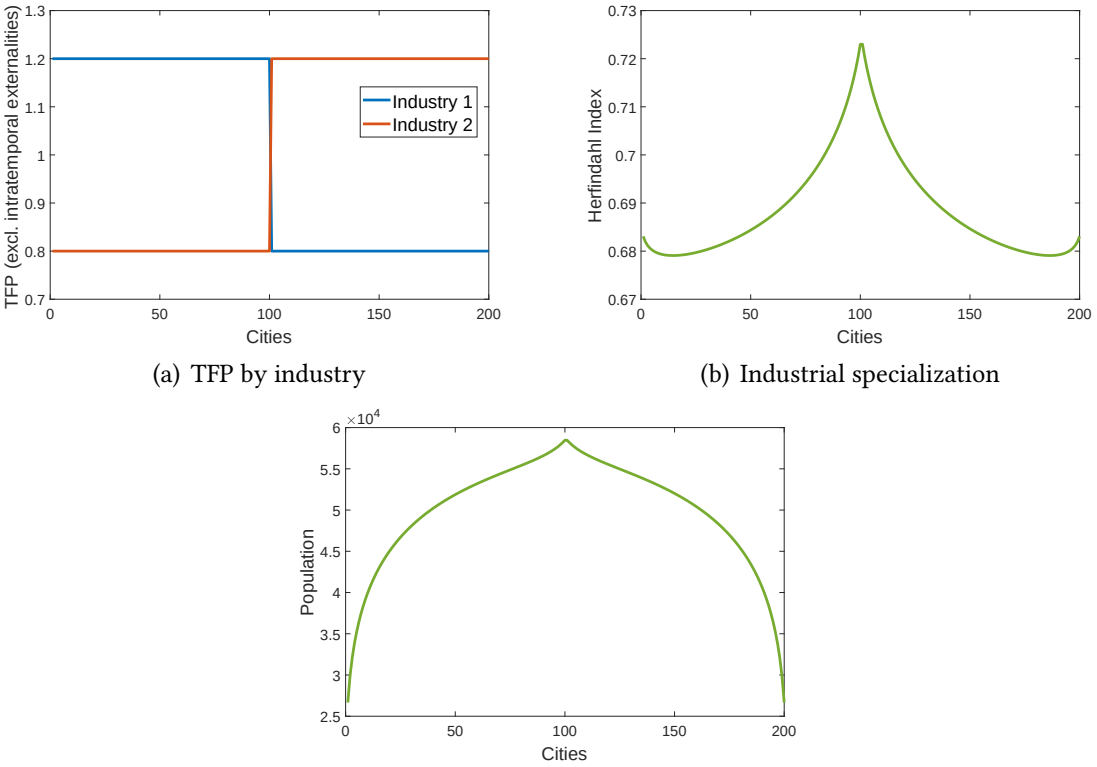
$$\tau_{icdt} = (1 + dist_{cd})^{\phi},$$

in both industries and time periods, where $dist_{cd}$ denotes the Euclidean distance between cities c and d. That is, trade is cheaper between cities that are geographically close.

We set the values of structural parameters to central values used in the literature ($\alpha = 0.06$, $\gamma = 0.65$, $\epsilon = 5$, $\phi = 0.25$, $\sigma = 4$). We set the total population to 10 million, which roughly equals the working population of England and Wales at the beginning of the 19th century. Finally, we set $\zeta_{ct} = 2$ for every city and time period for simplicity. This implies a land supply elasticity of one in each city.

The rise of (different) cities We first look at the patterns of specialization and the distribution of population in period 0, i.e., in the short run. The top right panel of Fig-

Figure 10. The linear economy under trade in period 0.



(a) TFP by industry

(b) Industrial specialization

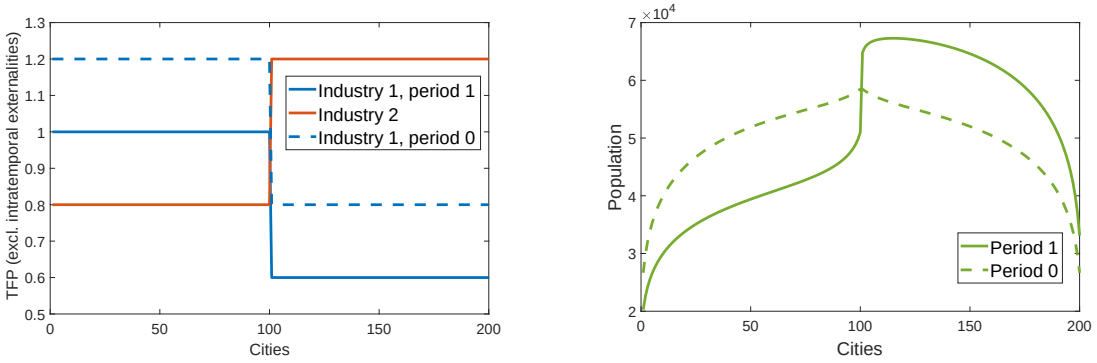
(c) Distribution of population.

Notes: The values for the structural parameters are set as follows: $\alpha = 0.06$; $\gamma = 0.65$; $\epsilon = 5$; $\phi = 0.25$; $\sigma = 4$; $\bar{L} = 10,000,000$; $\zeta_{ct} = 2$.

ure 10 shows how cities specialize under trade, as measured by their Herfindahl index across industries. One can see that cities near the center specialize more. Indeed, they are the ones with the best access to trade with other cities, and hence the largest room for specializing according to their comparative advantage. As the bottom panel of Figure 10 illustrates, population also reallocates towards these central cities as they benefit from trade through their increased specialization. This stands in stark contrast with the autarky scenario in which specialization and population are evenly distributed across cities, due to their symmetric fundamentals (see Figure C1 for an illustration). Thus, a short-run boom caused by trade favors cities in the center, which are able to gain from specialization and attract more people as a consequence.

The long-run dynamics and the death of (some) cities We study the distribution of economic activity in the long run (period 1) in two different cases. In the first case, we assume that industries differ in their aggregate productivity dynamics for exogenous reasons (outside the model). More precisely, we assume that the TFP of industry 1 decreases by 0.05 uniformly across cities, while the TFP of industry 2 stays the same as in

Figure 11. The linear economy in period 1 under differential industry trends.



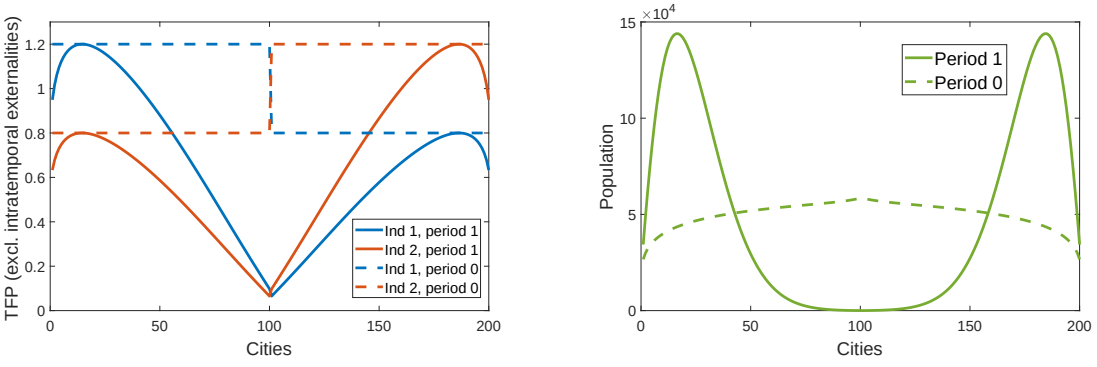
(a) TFP by industry (b) Distribution of population

Notes: values of structural parameters set to: $\alpha = 0.06$; $\gamma = 0.65$; $\epsilon = 5$; $\phi = 0.25$; $\sigma = 4$; $\bar{L} = 10,000,000$; $\zeta_{ct} = 2$.

period 0 (left panel of Figure 11). Such differential industry trends may be due to nationwide productivity trends associated with structural transformation, as in Ngai and Pissarides (2007), or increased international competition, as in Pierce and Schott (2016). We abstract from other sources of TFP evolution, such as dynamic externalities, by setting $f_i(\cdot) = 1$ in equation (11).

The right panel of Figure 11 shows that population reallocates from the West towards the East as a result of differential industry trends. This is not surprising as Western cities were the ones specializing in industry 1, hence they are the ones that suffer from declining TFP in this sector. Therefore, this case of the model can rationalize the boom and the subsequent bust of center-West cities, which initially specialized in the industry that starts declining after period 0.

Figure 12. The linear economy in period 1 under dynamic Jacobs externalities.



(a) TFP by industry (b) Distribution of population

Notes: values of structural parameters set to: $\alpha = 0.06$; $\gamma = 0.65$; $\epsilon = 5$; $\phi = 0.25$; $\sigma = 4$; $\bar{L} = 10,000,000$; $\zeta_{ct} = 2$.

In the second case, we abstract from differential industry trends but allow for dynamic agglomeration externalities. More precisely, we assume dynamic Jacobs externalities of the form

$$f_i(\cdot) = \left[\sum_j \left(\frac{L_{jc,t-1}}{L_{c,t-1}} \right)^2 \right]^{-1}$$

in equation (11). This formulation implies that cities less specialized in period 0 (those with a lower Herfindahl index, $\sum_j \left(\frac{L_{jc0}}{L_{c0}}\right)^2$) see faster TFP growth by period 1. The resulting TFP distribution is presented in the left panel of Figure 12.

As the right panel of Figure 12 illustrates, the long-term disadvantage of initially more specialized cities has the ability to reverse the hump-shaped population distribution of period 0. As a result, cities in the center see a period-1 bust after their period-0 boom.

How to tell apart differential industry trends from Jacobs externalities? Illustrating the model for the simple geography of this section highlights that differential industry trends and dynamic Jacobs externalities both have the ability to rationalize the death of specialized cities. However, they have different implications. In particular, differential industry trends imply that cities that hosted declining sectors suffer in the long-run. Dynamic Jacobs externalities, by contrast, imply that specialized cities suffer in the long run, no matter their industries. This is the insight that will allow us to disentangle these two mechanisms in Section 5.

5 Quantitative estimation

5.1 Taking the model to the data

To use the model for quantitative analysis, we take it to the data in this section. An advantage of the model's recursive structure is that it can be taken to the data in any period. Therefore, we take the model to twentieth-century ("long-run", period 1) data as richer data are available for this period than for the nineteenth century ("short-run", period 0).

Taking the model to the data consists of two steps. In the first step, we use the structure of the model to recover unobserved city-specific fundamentals that rationalize observed late-twentieth century (1971) data on employment by city-industry and wages by city. In the second step, we use the recovered TFP levels along with our instruments to estimate equation (15), i.e., the equation that drives the dynamic evolution of TFP in the model.

Recovering unobserved fundamentals In this step, we recover the distribution of city amenities $\bar{a}_c$ and city-industry productivities $\hat{T}_{ic1}$ that rationalize observed data on wages w_{c1} and employment by city-industry L_{ic1} in period 1. The following theorem states that there is a unique set of fundamentals that rationalize the data.

Theorem 1. *Given the values of structural parameters, trade costs in period 1, and data on wages w_{c1} and employment by city-industry L_{ic1}, one can uniquely recover TFP levels $\hat{T}_{ic1}$ (up to scale) and amenities relative to aggregate welfare, $\bar{a}_c/\bar{U}_1$.*

Proof. See Section C.2. □

Intuitively, the uniqueness result of Theorem 1 stems from the fact that the system of equations characterizing the equilibrium—equations (23) to (25) in Section C.1—can be inverted to recover the values of fundamentals.

Estimating the dynamic evolution of TFP Armed with the distribution of $\hat{T}_{ic1}$, we estimate the following equation characterizing the evolution of city-industry productivities:

$$\hat{T}_{ic1} = \bar{T}_{i1} L_{c0}^{\rho} \left[\sum_j \left(\frac{L_{jc0}}{L_{c0}} \right)^2 \right]^{-\iota} \varepsilon_{ic1} \tag{16}$$

Equation (16) is a special case of equation (15), i.e., the equation characterizing the evolution of TFP in the model. In particular, equation (16) restricts dynamic externalities, f_i, to be a function of two objects: city c's past population (L_{c0}) and past Herfindahl ($\sum_j \left(\frac{L_{jc0}}{L_{c0}}\right)^2$). This restriction is motivated by the reduced-form evidence of Section 3.3, which points to the importance of these two objects in explaining long-run city outcomes.

Equation (16) can be estimated by an empirical strategy that mirrors the one of Section 3.3. Specifically, one can use exogenous land fragmentation ζ_c and predicted Herfindahl $\hat{h}_c$ as instruments for late-nineteenth century (period-0) population and industrial specialization. Equation (16) corresponds to the second stage of this 2SLS procedure.

Note that we split the term T_{ic1} of equation (15) into two components in equation (16): an aggregate industry trend $\bar{T}_{i1}$ and a structural error term ε_{ic1}. In practice, the term $\bar{T}_{i1}$ can be estimated as an industry fixed effect in equation (16). Isolating these industry trends will allow us to disentangle them from the dynamic externalities, in line with the logic discussed in the end of Section 4.3.

5.2 Results [TBC]

6 Concluding remarks

Armed with two hundred years of data on British cities and exogenous variation in these cities' capacity to grow and specialize, we presented evidence that suggests intertemporal costs of the early success of these cities. Using our quantitative model, we suggest a mechanism: While many locations can indeed gain from being able to grow and specialize in response to structural transformation, this specialization comes at future costs as those locations fail to acquire the dynamic Jacobs externalities that sustain otherwise more diverse cities.

In future work, we will conduct the full estimation of the model using the data. We will then be in a position to use the model to conduct counterfactual analysis to ask, for example, how much the growth in trade translated into faster urbanization and city growth. We can also consider whether real aggregate gains arose, or whether the trade-induced changes caused merely a reorganization of activity across space. In that vein, we can also ask about the role of policy in a manner that is relevant for modern economies. Undoubtedly, attempts to replicate a 'Silicon Valley'-type success story in different parts of the world by focusing on growing sectors may be beneficial in the short-run. That said, the costs borne by formerly thriving hotspots of the industrial revolution that struggle to recover their previous advantages suggest potentially adverse long-run consequences of policy-driven specialization.

References

Allen, Robert C., *The British Industrial Revolution in Global Perspective*, Cambridge University Press, 2009.

_ , "The Interplay among Wages, Technology, and Globalization: The Labor Market and Inequality, 1620-2020," in "The Handbook of Historical Economics," Elsevier, 2021, pp. 795–824.

Allen, Treb and Dave Donaldson, "Persistence and Path Dependence in the Spatial Economy," Technical Report, National Bureau of Economic Research 2020.

_ **, Costas Arkolakis, and Xiangliang Li**, "On the Equilibrium Properties of Network Models with Heterogeneous Agents," Technical Report, National Bureau of Economic Research 2020.

Arribas-Bel, Daniel, M-À Garcia-López, and Elisabet Viladecans-Marsal, "Building (s and) cities: Delineating urban areas with a machine learning algorithm," *Journal of Urban Economics*, 2021, *125*, 103217.

Bairoch, Paul and Gary Goertz, "Factors of Urbanisation in the Nineteenth Century Developed Countries: A Descriptive and Econometric Analysis," Technical Report 1986.

Bellefon, Marie-Pierre De, Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin, "Delineating urban areas using building density," *Journal of Urban Economics*, 2021, *125*, 103226.

Berkes, Enrico, Ruben Gaetani, and Martí Mestieri, "Cities and Technological Waves," Technical Report, mimeo 2021.

Broadberry, S. N., B. M. S. Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen, *British Economic Growth, 1270-1870*, New York: Cambridge University Press, 2015.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and labor market dynamics: General equilibrium analysis of the China trade shock," *Econometrica*, 2019, *87* (3), 741–835.

Caprettini, Bruno and Hans-Joachim Voth, "Rage against the machines: Labor-saving technology and unrest in industrializing England," *American Economic Review: Insights*, 2020, *2* (3), 305–20.

Carlino, Gerald and William R Kerr, "Agglomeration and innovation," *Handbook of regional and urban economics*, 2015, *5*, 349–404.

Clark, Gregory, "'The Industrial Revolution', in Aghion, P. and Durlauf, S.N. (eds.)," *Handbook of Economic Growth*, 2014. Elsevier.

_ **and Anthony Clark**, "The Enclosure of English Common Lands, 1475-1839," *Journal of Economic History*, 2001, *61* (4), 1009–1036.

_ , **Kevin Hjortshøj O'Rourke, and Alan M. Taylor**, "The Growing Dependence of Britain on Trade during the Industrial Revolution," *Scandinavian Economic History Review*, May 2014, *62* (2), 109–136.

Coeurdacier, Nicholas, Florian Oswald, and Marc Teignier, "Structural Change, Land Use and Urban Expansion," Technical Report 2022.

Combes, Pierre-Philippe, Laurent Gobillon, and Yanos Zylberberg, "Urban economics in a historical perspective: Recovering data with machine learning," *Regional Science and Urban Economics*, 2022, *94*, 103711.

Crafts, Nicholas F. R., "British Industrialization in an International Context," *Journal of Interdisciplinary History*, 1989, *19* (3), 415.

Denman, Donald R., *Origins of Ownership: A Brief History of Land Ownership and Tenure in England from Earliest Times to the Modern Era*, George Allen & Unwin, 1958.

Duranton, Gilles, "Urban Evolutions: The Fast, the Slow, and the Still," *American Economic Review*, March 2007, *97* (1), 197–221.

_ **and Diego Puga**, "Nursery cities: Urban diversity, process innovation, and the life cycle of products," *American Economic Review*, 2001, *91* (5), 1454–1477.

Eckart, Wolfgang, "On the land assembly problem," *Journal of Urban Economics*, 1985, *18* (3), 364–378.

Eckert, Fabian and Michael Peters, "Spatial Structural Change," Technical Report, mimeo 2023.

Faggio, Giulia, Olmo Silva, and William C. Strange, "Heterogeneous agglomeration," *Review of Economics and Statistics*, 2017, *99* (1), 80–94.

Fajgelbaum, Pablo and Stephen Redding, "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914," Technical Report w20217 2021.

Franck, Raphaël and Oded Galor, "Flowers of Evil? Industrialization and Long-Run Development," *Journal of Monetary Economics*, 2021, *117*, 108–128.

Galor, Oded and Ömer Özak, "The agricultural origins of time preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

Glaeser, Edward L., "Reinventing Boston: 1630–2003," *Journal of Economic Geography*, 2005, *5*, 119–153.

Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer, "Growth in cities," *Journal of Political Economy*, 1992, *100* (6), 1126–1152.

Goose, Nigel, "Regions, 1700–1870," in Roderick Floud, Jane Humphries, and Paul Johnson, eds., *The Cambridge Economic History of Modern Britain*, second ed., Cambridge University Press, October 2014, pp. 149–177.

Hanlon, W. Walker, "The Rise of the Engineer: Inventing the Professional Inventor During the Industrial Revolution," Technical Report, mimeo 2021.

Hanlon, W Walker and Antonio Miscio, "Agglomeration: A long-run panel data approach," *Journal of Urban Economics*, 2017, *99*, 1–14.

Harari, Mariaflavia, "Cities in bad shape: Urban geometry in India," *American Economic Review*, 2020, *110* (8), 2377–2421.

Harley, C. Knick and N. F. R. Crafts, "Simulating the Two Views of the British Industrial Revolution," *The Journal of Economic History*, 2000, *60* (3), 819–841.

Heblich, Stephan, Alex Trew, and Yanos Zylberberg, "East-side story: Historical pollution and persistent neighborhood sorting," *Journal of Political Economy*, 2021, *129* (5), 1508–1552.

_ , Marlon Seror, Hao Xu, and Yanos Zylberberg, "Industrial clusters in the long run: evidence from Million-Rouble plants in China," Technical Report, CESifo Working Paper No. 7682 2019.

_ , Stephen J Redding, and Yanos Zylberberg, "The distributional consequences of trade: evidence from the repeal of the corn laws," Technical Report 2020.

Heldring, Leander, James A Robinson, and Sebastian Vollmer, "The economic effects of the English Parliamentary enclosures," Technical Report, National Bureau of Economic Research 2022.

Henderson, Vernon, Ari Kuncoro, and Matt Turner, "Industrial development in cities," *Journal of political economy*, 1995, *103* (5), 1067–1090.

Hennessy, Peter, *Never Again Britain 1945-1951*, London [u.a.: Penguin Books, 2006.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi, "Growth and Structural Transformation," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 2 of *Handbook of Economic Growth*, Elsevier, 2014, chapter 6, pp. 855–941.

Hills, Sally, Ryland Thomas, and Nicholas Dimsdale, "The UK Recession in Context — What Do Three Centuries of Data Tell Us?," *Bank of England Quarterly Bulletin*, 2010, p. 15.

Hoskins, William G., *The Making of the English Landscape*, London: Hodder and Stoughton, 1988.

Hudson, Pat, "Industrial Organisation and Structure," in Paul Johnson and Roderick Floud, eds., *The Cambridge Economic History of Modern Britain: Volume 1: Industrialisation, 1700–1860*, Vol. 1, Cambridge: Cambridge University Press, 2004, pp. 28–56.

Jacobs, Jane, *The Economy of Cities*, New York: Vintage, 1969.

Kelly, Morgan, Joel Mokyr, and Cormac O'Grada, "The Mechanics of the Industrial Revolution," Technical Report June 2020.

Matsuyama, Kiminori, "Agricultural productivity, comparative advantage, and economic growth," *Journal of Economic Theory*, 1992, *58* (2), 317–334.

Mendels, Franklin F., "Proto-Industrialization: The First Phase of the Industrialization Process," *The Journal of Economic History*, March 1972, *32* (1), 241–261.

Mokyr, Joel, *The Enlightened Economy: An Economic History of Britain 1700–1850*, Yale University Press, 2009.

_ , "Attitudes, Aptitudes, and the Roots of the Great Enrichment," in "The Handbook of Historical Economics," Elsevier, 2021, pp. 773–794.

Musson, Albert E., "Industrial Motive Power in the United Kingdom, 1800-70," *The Economic History Review*, 1976, *29* (3), 415–439.

Nagy, Dávid Krisztián, "Quantitative economic geography meets history: Questions, answers and challenges," *Regional Science and Urban Economics*, 2022, *94.*

Nagy, Dávid Krisztián, "Hinterlands, city formation and growth: Evidence from the US westward expansion," *Review of Economic Studies*, 2023.

Neeson, Jeanette M, *Commoners: common right, enclosure and social change in England, 1700-1820*, Cambridge University Press, 1996.

Ngai, L. Rachel and Christopher A. Pissarides, "Structural Change in a Multisector Model of Growth," *American Economic Review*, 2007, *97* (1), 429–443.

Nuvolari, Alessandro, Bart Verspagen, and Nick von Tunzelmann, "The Early Diffusion of the Steam Engine in Britain, 1700–1800: A Reappraisal," *Cliometrica*, October 2011, *5* (3), 291–321.

Ogilvie, Sheilagh, "Protoindustrialization," in "The New Palgrave Dictionary of Economics," London: Palgrave Macmillan UK, 2008, pp. 1–6.

Pierce, Justin R and Peter K Schott, "The surprisingly swift decline of US manufacturing employment," *American Economic Review*, 2016, *106* (7), 1632–62.

Redding, Stephen J. and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9*, 21–58.

Saiz, Albert, "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Satchell, Max and Dan Bogart and Leigh Shaw Taylor, "Parliamentary Enclosures in England, 1606–1902," Technical Report 2017. Data obtained on 2022-06-02.

Shaw-Taylor, L. and E. Anthony Wrigley, "Occupational structural and population change. Chapter 2 in Floud, R., Humphries, J. and Johnson P.," *The Cambridge Economic History of Modern Britain: Volume 1, Industrialisation 1700-1860*, 2014. Cambridge University Press.

Stokey, Nancy L., "A Quantitative Model of the British Industrial Revolution, 1780–1850," *Carnegie-Rochester Conference Series on Public Policy*, December 2001, *55* (1), 55–109.

Strange, William C., "Information, holdouts, and land assembly," *Journal of Urban Economics*, 1995, *38* (3), 317–332.

Trew, Alex, "Spatial Takeoff in the First Industrial Revolution," *Review of Economic Dynamics*, 2014, *17*, 707–25.

Trinder, Barrie, "Industrialising Towns 1700–1840," in Peter Clark, ed., *The Cambridge Urban History of Britain: Volume 2: 1540–1840*, Vol. 2 of *The Cambridge Urban History of Britain*, Cambridge: Cambridge University Press, 2000, pp. 805–830.

Turner, Michael, "Agricultural Productivity in England in the Eighteenth Century: Evidence from Crop Yields 1," *The Economic History Review*, 1982, *35* (4), 489–510.

Voigtländer, Nico and Hans-Joachim Voth, "The Three Horsemen of Riches: Plague, War, and Urbanization in Early Modern Europe," *The Review of Economic Studies*, April 2013, *80* (2), 774–811.

Wrigley, E. Anthony, "The PST System of Classifying Occupations," 2010.

ONLINE APPENDIX—not for publication

A An inspiration **41**

B Data appendix **42**

B.1 Data sources 42

B.2 Data construction 46

B.3 Descriptive statistics 48

B.4 City population and city growth: Additional evidence 50

B.5 Robustness 51

C Theory appendix **54**

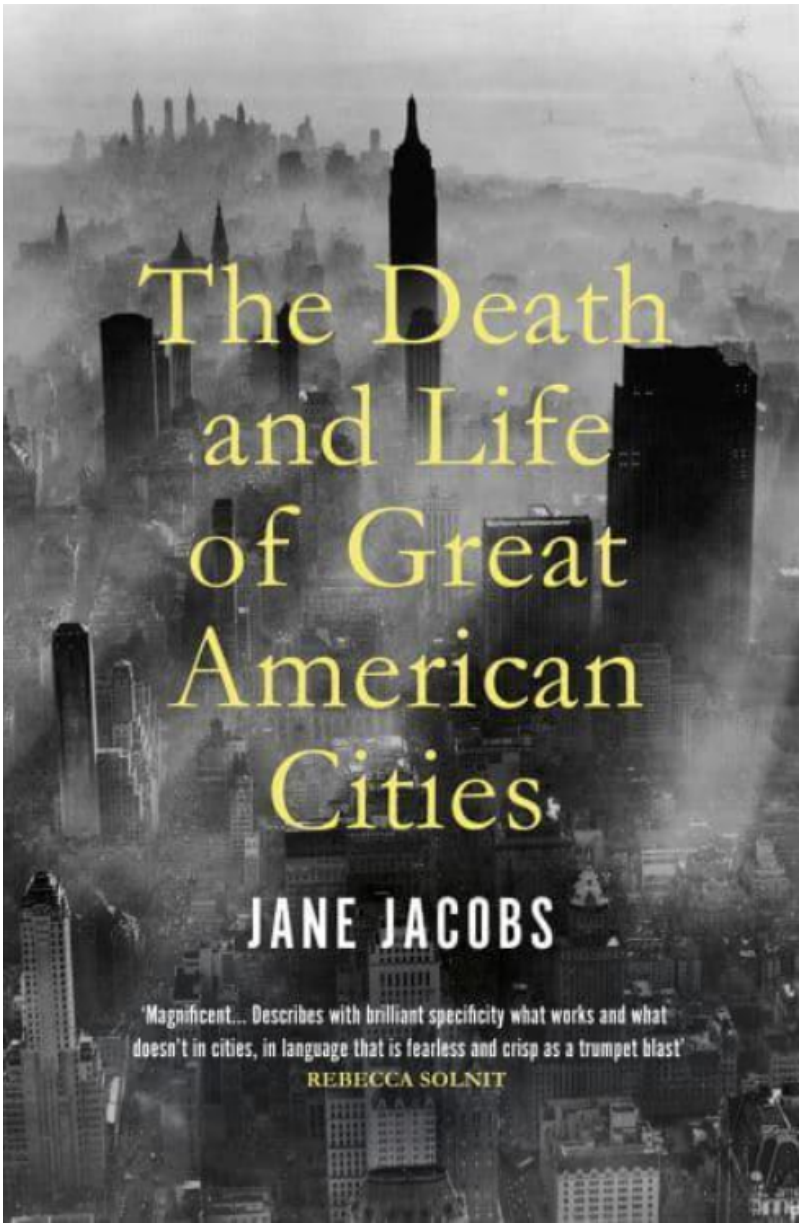
C.1 Derivation of equation (12) 54

C.2 Proof of Theorem 1 56

C.3 Complements to the linear economy 56

A An inspiration

Figure A1. The Death and Life of Great American Cities.



B Data appendix

This section provides complements to Section 2. More specifically, we detail our map digitization procedure, and we shed additional light on our land fragmentation algorithm.

B.1 Data sources

Recognizing built-up

To recognize built-up and other land use types in historical maps, we develop an object-based classification which is standard in remote sensing but less used in map digitization (Combes et al., 2022).

Figure B1. Recognizing built-up—superpixels.

Note: This Figure illustrates our approach based on superpixels (county map of Yorkshire; superpixels are delineated in green).

The first step consists in collecting and digitizing early county maps. A subsample of these maps was produced by the Ordnance Survey at the beginning of the nineteenth century; they are usually referred to as the Old Series 1-inch maps. We complement these maps with county-specific maps in the North of England.[16] We project these flat map

[16]Please find below the covered COUNTIES, with the map resolution, its date and its author(s): NORTHUMBERLAND, 1-inch map, 1769 (Andrew and Mostyn ARMSTRONG); DURHAM, 1-inch map 1768 (Thomas JEFFERYS and Andrew ARMSTRONG), 1-inch map 1820 (Christopher GREENWOOD);

tiles, which gives us an exhaustive coverage of England and Wales, albeit at different dates around 1800.

The second step produces superpixels of homogenous colors. There are several image segmentation algorithms, e.g., SLIC, Quickshift, Felzenszwalb or watershed.[17] We opt for "Quickshift", which allows for a grouping of pixels that is more flexible in their proximity in actual space (i.e., along the physical distance) versus the color space. We parametrize the algorithm by choosing the scale parameter, the maximum physical distance, and the relative weight between distance in the color-space and physical distance in order to best capture: built-up areas; smaller letters; and tree symbols within one superpixel. Figure B1 provides an example in Yorkshire, where we see how superpixels isolate letters, farms and the tree symbols that represent woods.

The third step consists in constructing variables that characterize these superpixels; these variables will be the input of a machine-learning classification algorithm. We can distinguish three types of variables: inner-superpixel variables capturing the color gradient within each superpixel, shape variables capturing the shape of superpixels, neighbor variables capturing the nature of neighboring superpixels. To capture the color and texture of superpixels, we select the following inner-superpixel variables: average R/G/B intensities, and within-standard deviation in R/G/B intensities (e.g., to capture stripes). To capture the shape of superpixels, we select the following shape variables: the compactness (minimum edge/maximum edge for the minimum bounding rectangle), area, perimeter and complexity of the superpixel (perimeter/ perimeter of the minimum bounding rectangle). To capture the nature of neighboring superpixels, we add their average R/G/B intensities to our set of predicting variables.

The fourth step consists in producing a training sample: i.e., a set of labeled superpixels (in a few categories: text, forest, sparse trees, built-up, etc.). We conduct this step using a simple manual labeling procedure whereby a window shows the outline of the superpixel and the underlying map and research assistants can select a land use category.

The fifth step trains a random forest algorithm on the training sample and predicts land use for all other non-labeled superpixels (Combes et al., 2022).

CUMBERLAND, 1/2-inch map, 1773 (Joseph HODSKINSON and Thomas DONALD), 1-inch map, 1823 (Christopher and John GREENWOOD); WESTMORLAND, 1-inch map, 1771 (Thomas JEFFERYS); YORKSHIRE, ca 4/5 inch, 1828 (Christopher GREENWOOD, et al); LANCASHIRE, not stated, 1768 (William YATES), 1 inch, 1818 (Christopher GREENWOOD, et al), 4/5 inch, 1830 (G. HENNET); CHESHIRE, 3/4 inch, 1830 (William SWIRE and W.F. HUTCHINGS), 5/4 inch, 1831 (Andrew BRYANT); DERBYSHIRE, 1-inch map 1767/1791 (Peter BURDETT); NOTTINGHAMSHIRE, not stated, 1794 (John CHAPMAN), not stated 1826 (Christopher and John GREENWOOD); LINCOLNSHIRE, ca. 1 inch, 1828 (A. Bryant), ca. 1 inch, 1830 (Christopher and John GREENWOOD).

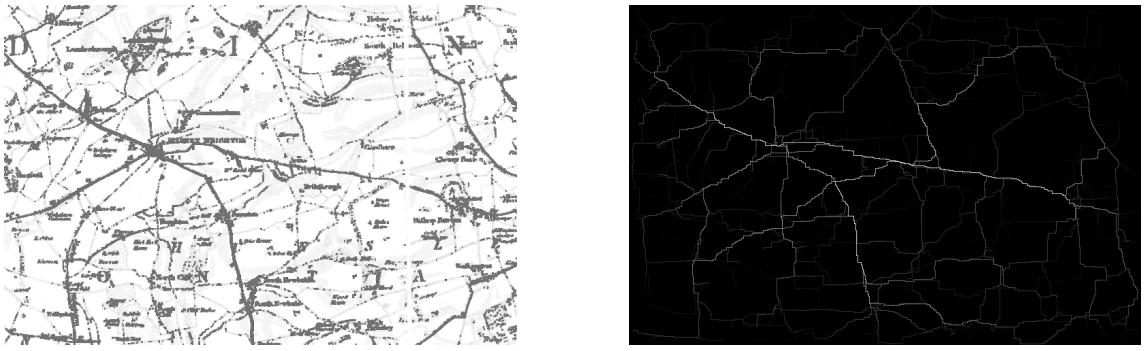
[17] See, e.g., https://scikit-image.org/docs/stable/auto_examples/segmentation/plot_segmentations.html.

Recognizing roads

To identify roads, we cannot use any standard procedure: most available procedures are designed to recognize straight lines,[18] and (i) our roads are not very straight and (ii) there are many other straight lines but road segments on county maps (e.g., letters or letter fragments).

Our approach relies on the nature of roads: roads are contiguous (impervious) areas which are designed to minimize travel cost between locations of interest. In practice, we can design a procedure to detect black lines, but we would then typically end up with many false positives (i.e., black lines that are not roads). Many of these black lines would, however, not fulfill the previous requirement: letters, gradient lines in mountainous terrain or farms are very inefficient patches of black pixels on which to travel between locations of interest. Others may be better suited, e.g., rivers, county boundaries, or railway tracks.

Figure B2. Recognizing roads.



(a) "Travel" cost (b) Paths

Note: Panel a displays the input of the least cost path procedure to detect roads on a historical county map. Panel b displays the simulated least cost paths drawn between random departure and arrival points.

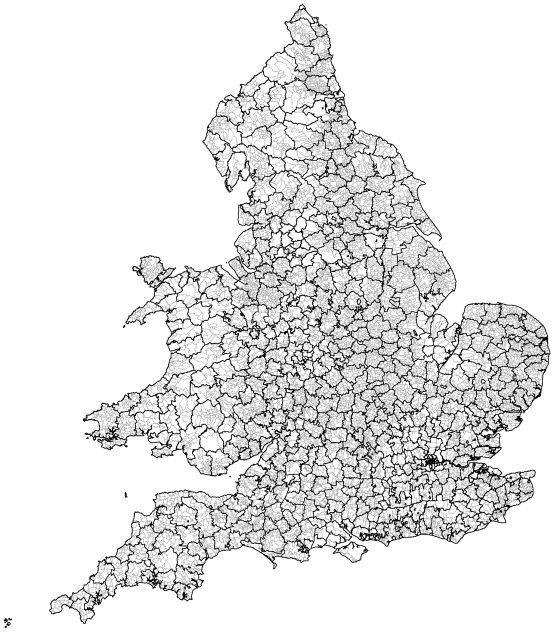
We exploit the previous discriminatory feature as follows. We first transform the color map into a gray "travel cost" matrix by (i) filtering out non-gray colors, (ii) thickening black areas to create continuous lines from dashed lines, (iii) simplifying images by interpolating across pixels, and (iv) transforming the gray intensity through a power function in order to best calibrate the "cost" of traveling across pixels of different black intensity (with travel cost over a perfectly black pixel being 1, and travel cost over a perfectly white pixel being as high as possible). The output is the left panel of Figure B2.

Second, we draw many random departure points from the subsample of black pixels on the image (as a proxy for the unobserved locations of interest) and we compute the

[18] See https://www.sciencedirect.com/topics/computer-science/hough-transforms.

least cost paths to all other pixels of the image for each draw. Roads are already more salient on this matrix. However, roads are designed to be used as "the" minimum cost path between locations. To better, we randomize many arrival points for each departure point and start drawing the actual minimum paths between departure and arrival pixels on a black image. The output is a distribution of many minimum cost paths highlighting the most traveled roads more than the others (right panel of Figure B2). As shown by Figure B2, this algorithm not only identifies roads very well, but also their respective importance within the transportation network.

Figure B3. Consistent parishes across England and Wales.



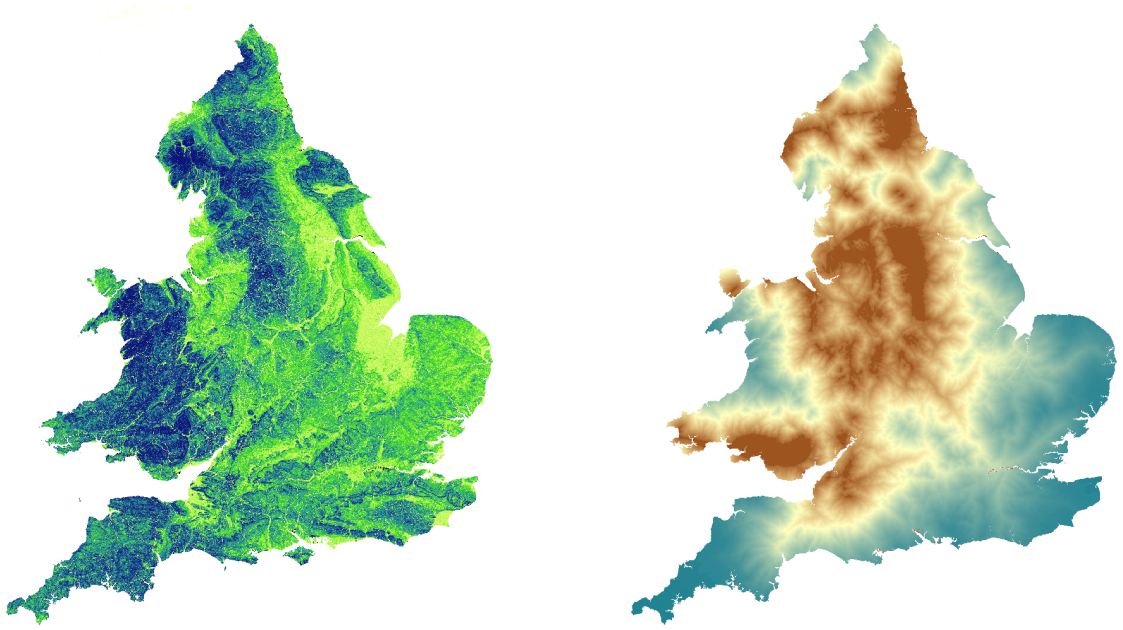
Notes: This Figure displays the output of the transitive closure algorithm implemented by the Cambridge Group for History of Population and Social Structure. Consistent mappable units based on parishes are displayed in gray; registration districts are displayed with black borders.

Consistent parishes and travel cost

One issue with census data and baptism records is that they are nested at the parish level, and parishes are regularly redefined, merged or split over the course of the nineteenth century. We thus apply an "envelope" algorithm which considers the union of the different parishes covering the same points over time. For instance, if a parish is split into two parishes in 1891, we would group the two sub-parishes from 1891 onward to create a consistent, unique parish from 1801 to the current day. This grouping is less relevant at the city level, as none of these re-compositions of lowest administrative units

significantly affect the allocation of administrative units across cities. The output of the procedure is shown in Figure B3, with about 11,500 consistent parishes across England and Wales.

Figure B4. Travel cost and distance to coal across England and Wales, as computed around 1817.



(a) Travel cost (b) Distance to coal

Notes: The left panel displays the raster of transport costs as calculated using the transport network at the beginning of the 19th century, and a penalization accounting for the local elevation gradient (yellow: low, green: medium, blue: high). The right panel displays the minimum travel time from the nearest coal field (red: low, blue/green: high).

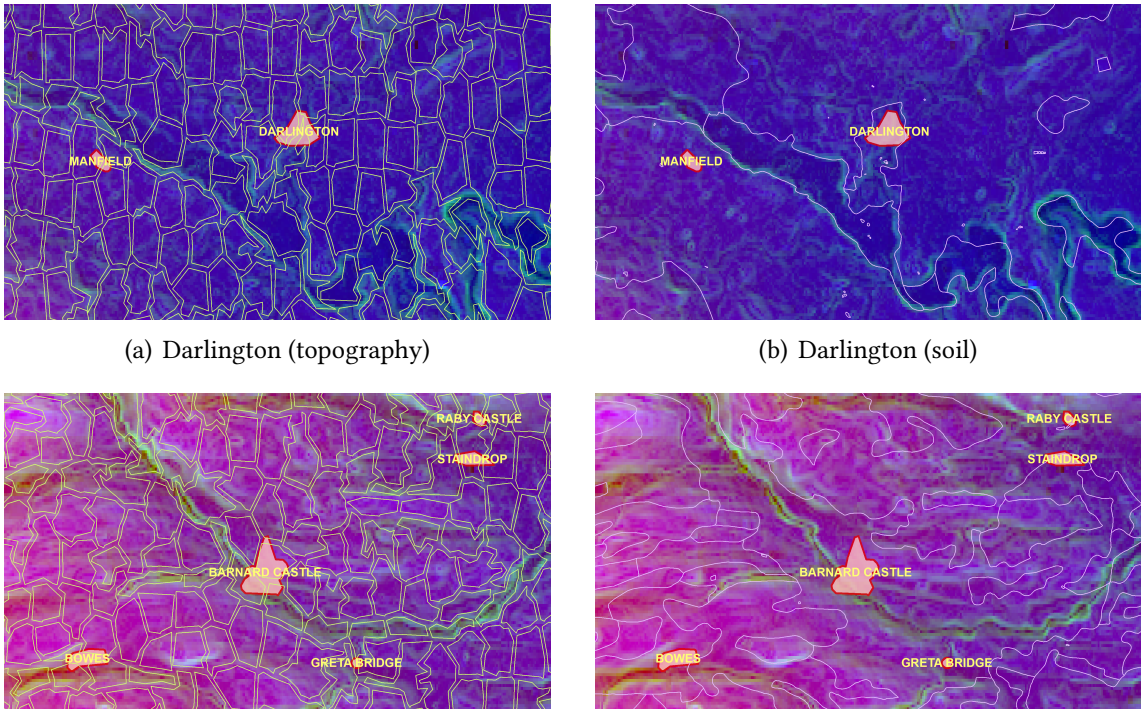
To compute trading costs over time, we gather roads, railways and waterways over the course of the nineteenth century (Transport, urbanization and economic development in England and Wales c.1670–1911, The Cambridge Group for the History of Population and Social Structure) and we apply a standard least cost path procedure across all cities of our sample, from our cities to the major trading ports and from our cities to major resources (see an illustration in Figure B4).

B.2 Data construction

Our land fragmentation algorithm consists of the following steps. In a first step, we extract fine-grained terrain characteristics including elevation, ruggedness, time-invariant soil attributes and water bodies from Google Earth Engine. We then store these characteristics in an "image", i.e., a 2D array where the value at each pixel (x, y) is a vector

of attributes $(a_1, \ldots, a_m)$—a standard color image is often stored as a 2D array where the value at each pixel is a triplet RGB. We then normalize all the different attributes to be between 0 (minimum) and 1 (maximum). The output is a multi-band raster covering England and Wales at a resolution of 30m.

Figure B5. Gradient breaks in topography and soil.



(a) Darlington (topography)

(b) Darlington (soil)

(c) Barnard Castle (topography)

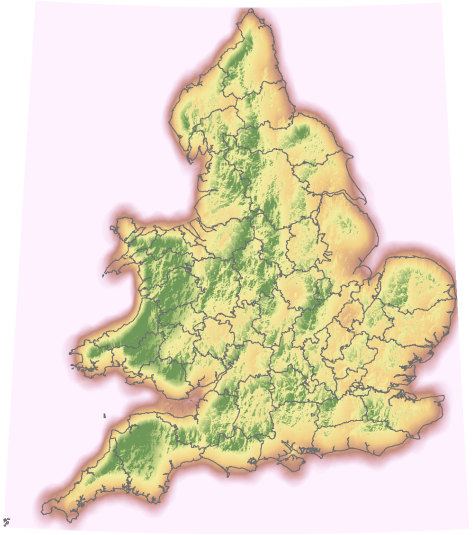
(d) Barnard Castle (soil)

Sources: topography (Google Earth Engine, 30m resolution), soil categories (National Soil Resources Institute).

In a second step, we use the same unsupervised segmentation algorithm as before, Quickshift, to isolate contiguous zones of the image that are homogeneous in their attributes. Applied to this peculiar setting, the algorithm relies on a standard "orthogonal" distance between the n-dimensional vectors that are stored in every pixel. The algorithm then maximizes a weighted sum of the two target distances within constituted superpixels, with a weight allocated to physical distance relative to the previously-defined "value distance." A superpixel with similar values is, here, a patch of land with homogeneous topography and soil characteristics: a typical agricultural parcel. For illustration, we show in Figure B5 the output of such an algorithm based on topography variable uniquely (left panel), and on soil characteristics uniquely (right panel).

In a third step, we construct the centroids of all superpixels, and we use these centroids to calculate the predicted density of farms at the fringe of each city—a measure of predicted land fragmentation. The variation in predicted land fragmentation is very

Figure B6. Predicted land fragmentation.



Note: This Figure displays an interpolated raster map of predicted land fragmentation (see Section 2.2). The color scheme is white (low) to red, yellow and green (high).

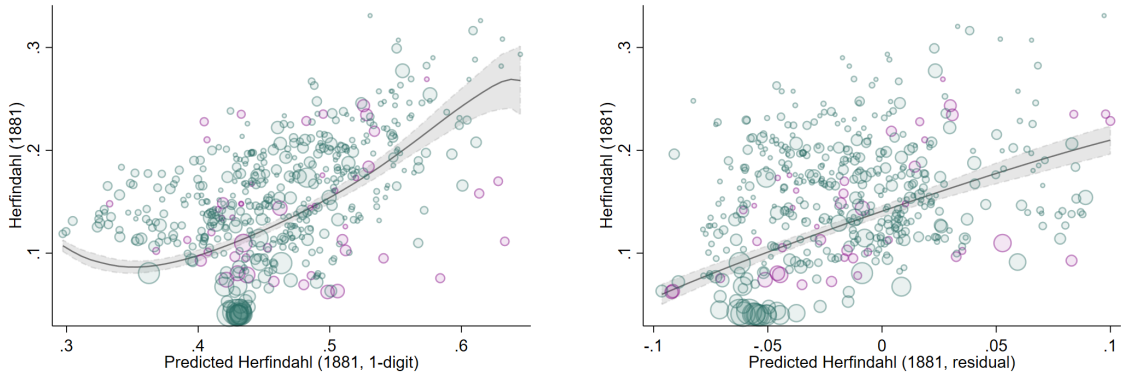
local; we do however see more aggregate patterns in land fragmentation due to overall topography (mountainous terrain in Central England or Wales) or soil composition (in Cornwall). We illustrate these aggregate patterns in Figure B6.

B.3 Descriptive statistics

In this section, we provide complements to Section 3.1. We first shed additional light on the role of 1-digit sectors in cities' changing industrial composition as opposed to the more granular variation within these 1-digit sectors. In Figure 9 of the main paper, we display the relationship between the 1881 Herfindahl index of specialization computed at the 2-digit level and a predicted measure, also computed at the 2-digit level. In Figure B7, we decompose this relationship into (i) the part explained by predicted industrial concentration at the 1-digit level (panel a) and (ii) the residual of the 1-digit prediction, thus exploiting within-sector variation only (panel b). We can see that the relationship remains strong in panel b; the slope is about 60% to that of Figure 9.

Next, we illustrate the correlation between settlement characteristics at the beginning of the nineteenth century and predicted land fragmentation in these settlements' immediate fringe in Figure B8. We find that settlements with above- (red) and below-median (blue) predicted land fragmentation are very similar in population density and industrial specialization.

Figure B7. The role of the initial industrial mix—robustness checks.

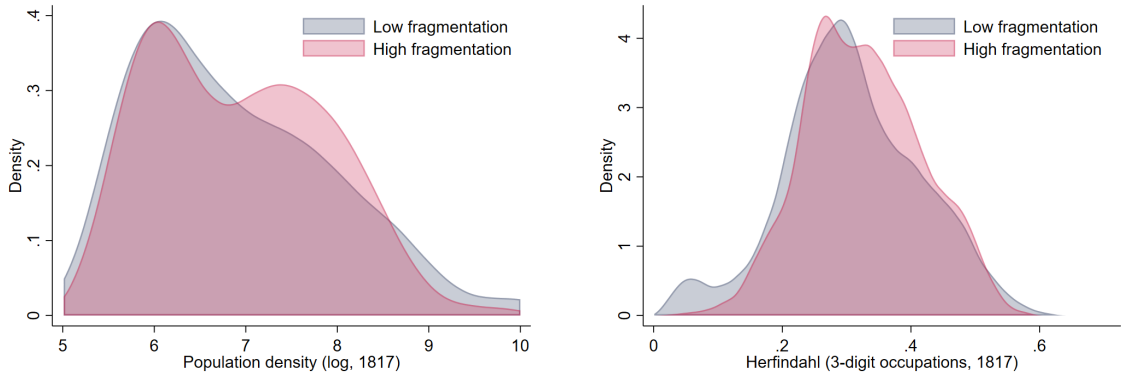


(a) Specialization (1881, 1-digit)

(b) Specialization (1881, residual)

Notes: The left panel displays the measure of predicted industrial concentration computed at the 1-digit level against the actual industrial concentration. The top decile of growing settlements are highlighted in purple. The line is a locally weighted regression on all observations. The right panel repeats the same exercise with a residualized measure of predicted industrial concentration that is (i) computed at the 2-digit level and (ii) cleaned for the role of the previous 1-digit measure of predicted industrial concentration. The slope is about 60% to that of Figure 9 (panel b).

Figure B8. A balance test for urban settlements with different land fragmentation in their immediate fringe.



(a) Population density

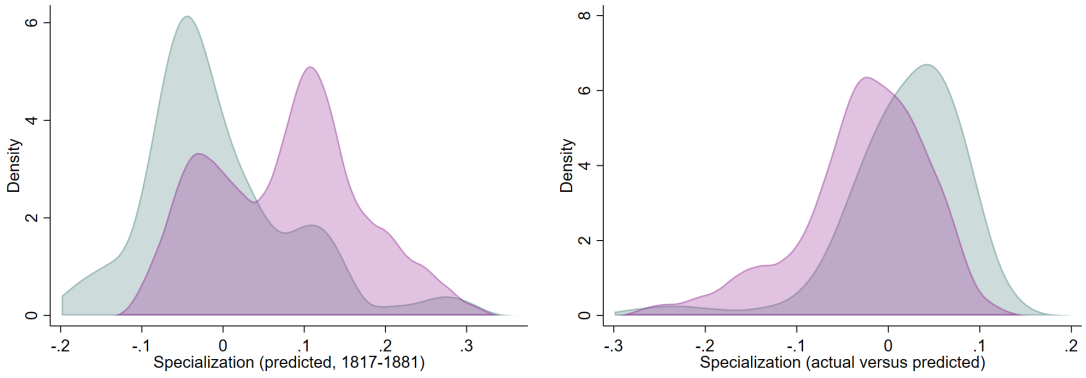
(b) Industrial specialization (Herfindahl index)

Notes: This Figure displays the distribution of population density (left panel) and industrial specialization (calculated as a Herfindahl index across 3-digit occupations, right panel) for cities with above- (red) and below-median (blue) predicted land fragmentation in their immediate fringe.

In Section 3.1, we also show that fast-growing cities become more specialized. We better qualify this observation in Figure B9 where we decompose specialization h_c (as shown in Figure 8) into predicted specialization $\hat{h}_c$ (see Section 2.2) and residual specialization $h_c - \hat{h}_c$. We then plot the distribution of these two objects for the top quartile in terms of employment growth during 1817–1881 (in purple) and the rest (in teal). We see that the specialization of fast-growing cities is entirely predicted from their initial

industry mix: fast-growing cities are initially specialized across a few key sectors that flourish over the course of the nineteenth century.

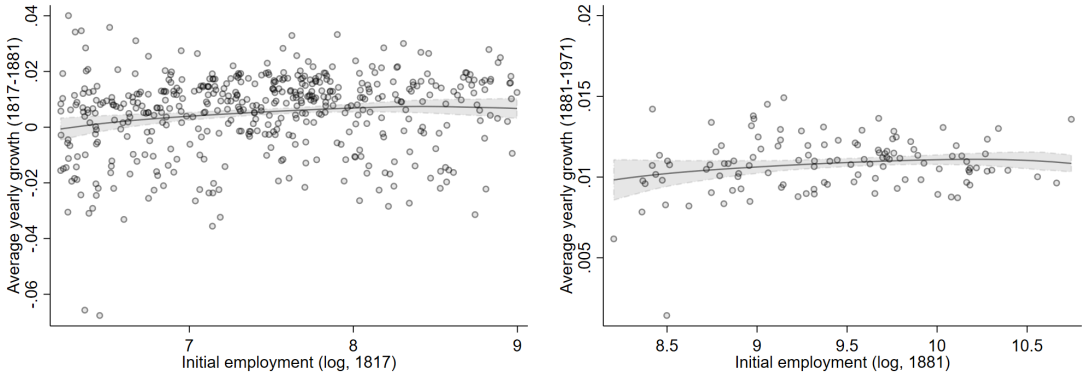
Figure B9. Specialization of cities—prediction and residual Herfindahl.



(a) Predicted specialization (b) Residual specialization

Notes: This Figure shows the distribution of predicted specialization $\hat{h}_c$ (see Section 2.2, left panel) and residual specialization $h_c - \hat{h}_c$ (right panel). The top decile of growing settlements are highlighted in purple.

Figure B10. Gibrat's law.



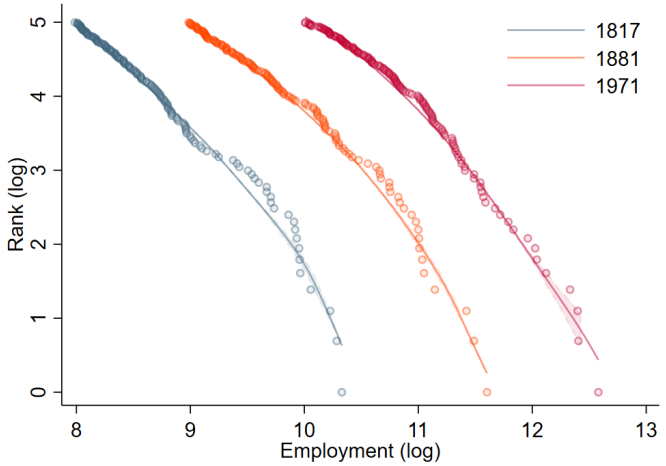
(a) Gibrat's law, 1817–1881 (b) Gibrat's law, 1881–1971

Notes: This Figure shows the relationship between cities' (log) initial employment and subsequent yearly growth. Panel (a) conducts the analysis for the period 1817–1881, while panel (b) conducts it for the period 1881–1971.

B.4 City population and city growth: Additional evidence

This section presents additional evidence about the distribution of city populations and growth over time. In Figure B10, we display the relationship between cities' average yearly growth and initial employment. Panel (a) shows this relationship in the nineteenth century. As expected, initially large cities grow faster during this period. Panel (b)

Figure B11. Zipf's law.



Notes: This Figure shows the relationship between cities' (log) rank in 1817 (blue), 1881 (orange) and 1971 (purple) and their (log) employment.

repeats the analysis for the twentieth century. Over this period, the relationship becomes flat, even slightly negative: that is, large cities lose their advantage in terms of growth. This is in line with the empirical findings of Section 3.3: larger population does not confer an advantage in terms of long-run growth. It is also in line with the urban literature documenting a typically flat relationship between city size and growth (Gibrat's Law).

We also provide support for another empirical regularity characterizing the distribution of city size: the Zipf law relating the rank of cities to their size. In Figure B11, we plot the relationship between the (log) rank and the (log) employment across our cities in 1817 (blue), 1881 (orange) and 1971 (purple). One can see that the relationship is close to linear in all years; the slope is, however, not close to -1, but is between -2 and -3.

B.5 Robustness

In this section, we provide a sensitivity analysis for our main finding of Section 3.3. First, we provide evidence that land fragmentation affects urban development even when controlling for competing mechanisms, more specifically: general agricultural productivity (Coeurdacier et al., 2022); returns to agriculture and the exposure to the repeal of the Corn Laws (Heblich et al., 2020); and the role of local institutions through the prevalence of parliamentary enclosures (Heldring et al., 2022).

Table B1 reports a sensitivity analysis around our baseline specification (Table 3, column (2), replicated in column (1) of Table B1). In column (2), we control for the average Caloric Suitability Index within a 3-kilometer buffer around the initial city boundaries

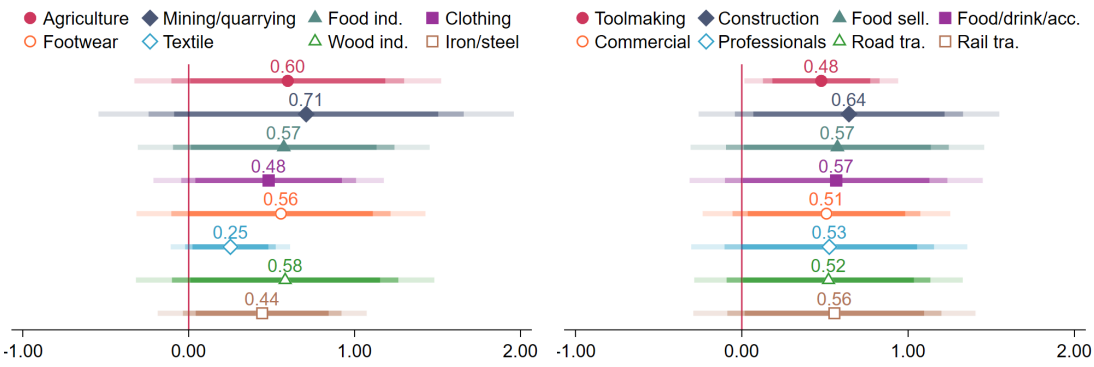
Table B1. The long-run effect of city population and specialization—robustness checks.

	Unskilled workers (1971)				
	(1)	(2)	(3)	(4)	(5)
Herfindahl (1881)	0.779	0.754	0.788	0.621	0.809
	(0.275)	(0.283)	(0.309)	(0.253)	(0.407)
Population (1881)	0.017	0.020	0.024	0.021	0.032
	(0.011)	(0.011)	(0.012)	(0.010)	(0.018)
Observations	285	259	259	259	239
Additional controls	—	Agricultural productivity	Exposure to Corn Laws	Most common crops' yields	Incidence of enclosures

Notes: The set of baseline controls include initial employment and Herfindahl in 1817, and: slope, elevation, bulk density, carbon content, latitude/longitude, travel time to coal, travel time to London, distance to/density of roads (1830), distance to/density of stations (1846), distance to waterways, distance to the shore, share of arable, suitability to grow wheat and grass, the "shift-share" predictor of urban employment growth ($\hat{g}_c$) between 1817 and 1881 defined in Section 2.2, and the same "shift-share" predictor constructed between 1881 and 1971.

(Galor and Özak, 2016) and for the potential yield for wheat (GAEZ: high input, rain-fed). In column (3), we control for the local share of arable in agricultural production in a broader neighborhood around the city, as captured by the Tithe maps in 1836. In column (4), we control for a decomposition of agricultural acreage between wheat, barley, oats, potato, peas, bean, rape, and turnip from the agricultural census in 1801. In column (5), we control for the local incidence of parliamentary enclosures (Clark and Clark, 2001; Heldring et al., 2022; Satchell, Max and Dan Bogart and Leigh Shaw Taylor, 2017). While we used a 3-kilometer buffer for all measures so far, we extend the buffer to a 10-kilometer buffer for the latter measure.

Figure B12. The long-run effect of city specialization—dropping industries one by one.



Notes: This Figure shows the estimated effect of 1881 industrial specialization on the 1971 share of unskilled in specifications in which we drop each of the 16 major nineteenth-century industries one by one.

Next, we examine whether our main empirical finding is driven by any one of the major nineteenth-century industries. To this end, we re-run the baseline specification of

Table 3, column (2), after dropping each of the 16 major industries, one by one. Figure B12 shows that the estimated effect of specialization on the long-run share of unskilled population remains positive in all of these specifications. This suggests that none of the major industries are driving our headline empirical result.

C Theory appendix

C.1 Derivation of equation (12)

Given that workers are freely mobile across industries, nominal wages equalize across industries in each city:

$$w_{ict} = w_{ct}.$$

Plugging this result into equation (7), we obtain total land rents in city c as

$$R_{ct}L_{ct} = r_{ct}H_{ct} = \frac{1-\gamma}{\gamma}w_{ct}L_{ct},$$

from which

$$w_{ct} + R_{ct} = \frac{1}{\gamma}w_{ct}. \tag{17}$$

Also, from equation (6), we get

$$r_{ct} = \left(\frac{1-\gamma}{\gamma}\right)^{1/\zeta_{ct}} w_{ct}^{1/\zeta_{ct}} L_{ct}^{1/\zeta_{ct}}. \tag{18}$$

By perfect competition, the factory gate price of each variety c in industry i becomes equal to its marginal cost of production in equilibrium:

$$p_{ict} = T_{ict}^{-1} w_{ct}^{\gamma} r_{ct}^{1-\gamma} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{1-\gamma}{\zeta_{ct}}} T_{ict}^{-1} L_{ct}^{\frac{1-\gamma}{\zeta_{ct}}} w_{ct}^{\gamma+\frac{1-\gamma}{\zeta_{ct}}}, \tag{19}$$

where we used equation (18). As a result, we can write the price index of industry i, equation (9), as

$$P_{idt} = \left[\sum_{c=1}^{C} \left(\frac{1-\gamma}{\gamma}\right)^{-\frac{1-\gamma}{\zeta_{ct}}(\epsilon-1)} T_{ict}^{\epsilon-1} L_{ct}^{-\frac{1-\gamma}{\zeta_{ct}}(\epsilon-1)} w_{ct}^{-\left(\gamma+\frac{1-\gamma}{\zeta_{ct}}\right)(\epsilon-1)} \tau_{icdt}^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}}, \tag{20}$$

and market clearing condition (8) as

$$w_{ct}L_{ict} = \left(\frac{1-\gamma}{\gamma}\right)^{-\frac{1-\gamma}{\zeta_{ct}}(\epsilon-1)} T_{ict}^{\epsilon-1} L_{ct}^{-\frac{1-\gamma}{\zeta_{ct}}(\epsilon-1)} w_{ct}^{-\left(\gamma+\frac{1-\gamma}{\zeta_{ct}}\right)(\epsilon-1)} \sum_{d=1}^{C} P_{dt}^{\sigma-1} P_{idt}^{\epsilon-\sigma} w_{dt} L_{dt} \tau_{icdt}^{1-\epsilon}, \tag{21}$$

where we also used equation (17).

By the Fréchet distribution of idiosyncratic city tastes, the probability that a worker

chooses city c equals

$$Pr\left[U_{ct}^{m} \geq U_{dt}^{m}\ \forall d\right] = \frac{\left(\bar{a}_c \frac{w_{ct}+R_{ct}}{P_{ct}} L_{ct}^{-\lambda}\right)^{1/\eta}}{\sum_{d=1}^{C}\left(\bar{a}_d \frac{w_{dt}+R_{dt}}{P_{dt}} L_{dt}^{-\lambda}\right)^{1/\eta}} = \frac{\left(\frac{1}{\gamma}\bar{a}_c \frac{w_{ct}}{P_{ct}} L_{ct}^{-\lambda}\right)^{1/\eta}}{\sum_{d=1}^{C}\left(\frac{1}{\gamma}\bar{a}_d \frac{w_{dt}}{P_{dt}} L_{dt}^{-\lambda}\right)^{1/\eta}}.$$

In equilibrium, the fraction of workers choosing to live in c becomes equal to this probability:

$$\frac{L_{ct}}{\bar{L}} = \frac{\left(\frac{1}{\gamma}\bar{a}_c \frac{w_{ct}}{P_{ct}} L_{ct}^{-\lambda}\right)^{1/\eta}}{\sum_{d=1}^{C}\left(\frac{1}{\gamma}\bar{a}_d \frac{w_{dt}}{P_{dt}} L_{dt}^{-\lambda}\right)^{1/\eta}}$$

from which

$$P_{ct} = \left(\gamma \bar{U}_t\right)^{-1} \bar{a}_c w_{ct} L_{ct}^{-(\lambda+\eta)} \tag{22}$$

where

$$\bar{U}_t = \left[\frac{\sum_{d=1}^{C}\left(\frac{1}{\gamma}\bar{a}_d \frac{w_{dt}}{P_{dt}} L_{dt}^{-\lambda}\right)^{1/\eta}}{\bar{L}}\right]^{\eta}.$$

Plugging this result into equations (10) and (21) and rearranging equation (20), we obtain the system of equations

$$P_{ict}^{1-\epsilon} = \sum_{d=1}^{C}\left(\frac{1-\gamma}{\gamma}\right)^{-\frac{1-\gamma}{\zeta_{dt}}(\epsilon-1)} \hat{T}_{idt}^{\epsilon-1} L_{dt}^{-\left(\frac{1-\gamma}{\zeta_{dt}}-\alpha\right)(\epsilon-1)} w_{dt}^{-\left(\gamma+\frac{1-\gamma}{\zeta_{dt}}\right)(\epsilon-1)} \tau_{idct}^{1-\epsilon} \tag{23}$$

$$\bar{a}_c^{1-\sigma} w_{ct}^{1-\sigma} L_{ct}^{(\lambda+\eta)(\sigma-1)} = \left(\gamma \bar{U}_t\right)^{1-\sigma} \sum_{i=1}^{I} P_{ict}^{1-\sigma} \tag{24}$$

$$\begin{aligned}&\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1-\gamma}{\zeta_{ct}}(\epsilon-1)} w_{ct}^{1+\left(\gamma+\frac{1-\gamma}{\zeta_{ct}}\right)(\epsilon-1)} L_{ct}^{1+\left(\frac{1-\gamma}{\zeta_{ct}}-\alpha\right)(\epsilon-1)} = \\ &\left(\gamma \bar{U}_t\right)^{1-\sigma} \sum_{i=1}^{I}\sum_{d=1}^{C} \hat{T}_{ict}^{\epsilon-1} P_{idt}^{\epsilon-\sigma} \bar{a}_d^{\sigma-1} w_{dt}^{\sigma} L_{dt}^{1-(\lambda+\eta)(\sigma-1)} \tau_{icdt}^{1-\epsilon}\end{aligned} \tag{25}$$

where

$$\hat{T}_{ict} = T_{ict} f_i\left(L_{c,t-1}, \left\{L_{jc,t-1}\right\}_{j\in I}\right),$$

is the part of TFP that is exogenous in period t. Applying the change in variables in equation (14) and recalling $w_{ict} = w_{ct}$, equation (12) immediately follows from equations (23), (24) and (25).

C.2 Proof of Theorem 1

Rearranging the period-1 version of equations (23) to (25) yields

$$P_{ic1}^{1-\epsilon} = \sum_{d=1}^{C} \left(\frac{1-\gamma}{\gamma}\right)^{-\frac{1-\gamma}{\xi_{d1}}(\epsilon-1)} \hat{T}_{id1}^{\epsilon-1} L_{d1}^{-\left(\frac{1-\gamma}{\xi_{d1}}-\alpha\right)(\epsilon-1)} w_{d1}^{-\left(\gamma+\frac{1-\gamma}{\xi_{d1}}\right)(\epsilon-1)} \tau_{idc1}^{1-\epsilon}$$

$$\left(\frac{\bar{a}_c}{\bar{U}_1}\right)^{1-\sigma} w_{c1}^{1-\sigma} L_{c1}^{(\lambda+\eta)(\sigma-1)} = \gamma^{1-\sigma} \sum_{i=1}^{I} P_{ic1}^{1-\sigma}$$

$$\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1-\gamma}{\xi_{c1}}(\epsilon-1)} w_{c1}^{1+\left(\gamma+\frac{1-\gamma}{\xi_{c1}}\right)(\epsilon-1)} L_{c1}^{\left(\frac{1-\gamma}{\xi_{c1}}-\alpha\right)(\epsilon-1)} \hat{T}_{ic1}^{1-\epsilon} L_{ic1} =$$
$$\gamma^{1-\sigma} \sum_{d=1}^{C} P_{id1}^{\epsilon-\sigma} \left(\frac{\bar{a}_d}{\bar{U}_1}\right)^{\sigma-1} w_{d1}^{\sigma} L_{d1}^{1-(\lambda+\eta)(\sigma-1)} \tau_{icd1}^{1-\epsilon}$$

from which we obtain the following system of $3IC$ equations:

$$\hat{x}_{ic}^1 = \sum_{j=1}^{I} \sum_{d=1}^{C} \left(\hat{x}_{jd}^3\right)^{-1} \hat{K}_{icjd}^1$$
$$\hat{x}_{ic}^2 = \sum_{j=1}^{I} \sum_{d=1}^{C} \left(\hat{x}_{jd}^1\right)^{\frac{\sigma-1}{\epsilon-1}} \hat{K}_{icjd}^2$$
$$\hat{x}_{ic}^3 = \sum_{j=1}^{I} \sum_{d=1}^{C} \left(\hat{x}_{jd}^1\right)^{-\frac{\epsilon-\sigma}{\epsilon-1}} \left(\hat{x}_{jd}^2\right)^{-1} \hat{K}_{icjd}^3$$

where $\hat{x}_{ic}^1 = P_{ic1}^{1-\epsilon}$, $\hat{x}_{ic}^2 = \left(\frac{\bar{a}_c}{\bar{U}_1}\right)^{1-\sigma}$, and $\hat{x}_{ic}^3 = \hat{T}_{ic1}^{1-\epsilon}$.

The solution to this system exists and is unique if the largest eigenvalue of matrix $\mathbf{A}$ is less or equal to one in absolute value (Allen et al., 2020), where $\mathbf{A}$ is

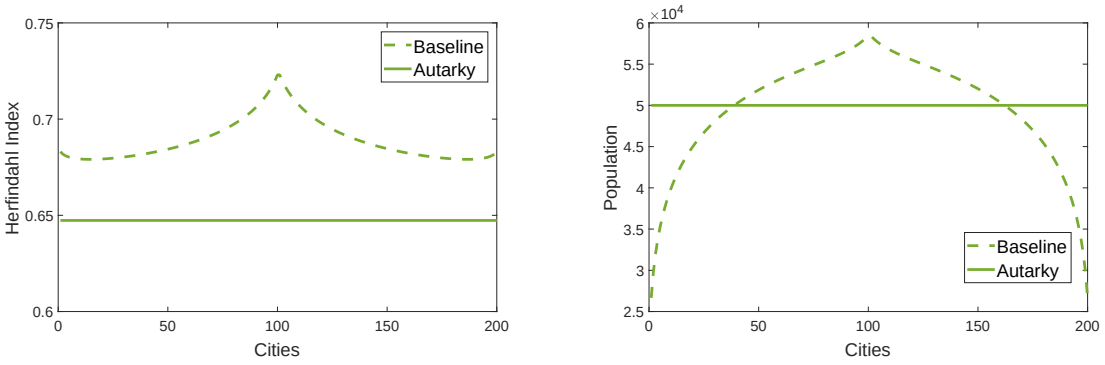
$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ \left|\frac{\sigma-1}{\epsilon-1}\right| & 0 & 0 \\ \left|\frac{\epsilon-\sigma}{\epsilon-1}\right| & 1 & 0 \end{bmatrix}.$$

One can easily verify that, under the assumption $\epsilon > \sigma > 1$, the largest eigenvalue of $\mathbf{A}$ is equal to 1.

C.3 Complements to the linear economy

Figure C1 presents the spatial distribution of industrial specialization and population in a world in which cities are in autarky. Figure C2 presents the same outcomes in a world in which the land supply elasticity of Western cities is decreased to zero.

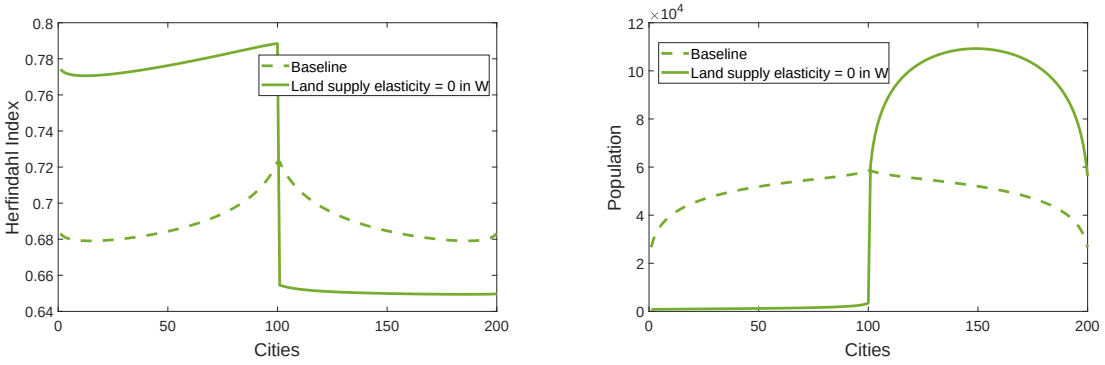
Figure C1. The linear economy under autarky in period 0.



(a) Industrial specialization. (b) Distribution of population.

Notes: The values for the structural parameters are set as follows: $\alpha = 0.06$; $\gamma = 0.65$; $\epsilon = 5$; $\phi = 0.25$; $\sigma = 4$; $\bar{L} = 10,000,000$; $\zeta_{ct} = 2$.

Figure C2. The linear economy in period 0 with heterogeneous land supply.



(a) Industrial specialization. (b) Distribution of population.

Notes: The values for the structural parameters are set as follows: $\alpha = 0.06$; $\gamma = 0.65$; $\epsilon = 5$; $\phi = 0.25$; $\sigma = 4$; $\bar{L} = 10,000,000$.