

Prime Locations

Gabriel M. Ahlfeldt, Thilo N. H. Albers, Kristian Behrens

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Prime Locations

Abstract

We harness big data to detect prime locations—large clusters of knowledge-based tradable services—in 125 global cities and track changes in the within-city geography of prime service jobs over a century. Historically smaller cities that did not develop early public transit networks are less concentrated today and have prime locations farther away from their historic cores. We rationalize these findings in an agent-based model that features extreme agglomeration, multiple equilibria, and path dependence. Both city size and public transit networks anchor city structure. Exploiting major disasters and using a novel instrument—subway potential—we provide causal evidence for these mechanisms and disentangle size- from transport network effects.

JEL-Codes: R380, R520, R580.

Keywords: prime services, internal city structure, agent-based model, multiple equilibria and path dependence, transport networks.

Gabriel M. Ahlfeldt
London School of Economics / United Kingdom
g.ahlfeldt@lse.ac.uk

Thilo N. H. Albers
Humboldt University Berlin / Germany
thilo.nils.hendrik.albers@hu-berlin.de

Kristian Behrens
Université du Québec à Montréal / Canada
behrens.kristian@uqam.ca

November 18, 2020

We gratefully acknowledge CEP, SERC and the LSE research committee for funding. We are indebted to Christian Hilber and Felix Weinhardt without whose input this project would not have come about. We thank seminar and conference participants in Barcelona (Autonoma), London (SERC), Düsseldorf (EGIT), Toronto (Sauder-UBC workshop), Montreal (UQAM-Bank of Canada workshop), Saint Petersburg (HSE), the Urban Economics Association Virtual Meeting, and especially Felipe Carozzi, Paul Cheshire, Gabriel Felbermayr, Steve Gibbons, Stephan Heblich, Vernon Henderson, Henry Overman, Olmo Silva, Daniel Sturm, Jens Südekum, Nick Tsivanidis, and Niko Wolf for comments and suggestions. We also thank Marco Gonzalez-Navarro and Matt Turner for sharing their subway data with us at an early stage of the project. Katia Del-Frari, Peter Jeffrey, Kalle Kappner, Stephen Law, Sascha Moebius, Alice Neht, Tiantong Song, Timo Stieglitz, and Sevrin Waights provided excellent research assistance. Behrens gratefully acknowledges financial support from: the CRC Program of the Social Sciences and Humanities Research Council of Canada for the funding of the Canada Research Chair in Regional Impacts of Globalization; and SSHRC Insight Grant # 435-2016-1246. This article was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project '5-100'. Albers gratefully acknowledges the support of the CRC TRR 190 (DFG). The usual disclaimer applies.

1 Introduction

The importance of prime services—knowledge-based tradable services such as finance, insurance, real estate, and professional services—has grown tremendously over the last century. Their share in US GDP is more than twice that of manufacturing. They are the economic engines of the world’s Top-100 “global cities” which generate close to 40% of Earth’s GDP (Dobbs et al., 2011). Within global cities, prime service activity is concentrated ‘on the head of a pin’, with hundreds of thousands of high-paying jobs within walkable areas such as Midtown in New York, the City of London, the Central District in Hong Kong, or Marunouchi in Tokyo. Our data for 125 global cities reveal that about 38% of their prime service employment is clustered within about 0.7% of their developable area. These hyper-productive and dense clusters of wealth creation—prime locations—dominate business centers and are the powerhouses of the global economy. Understanding the factors that determine their emergence, location, resilience, and change is of great policy interest, especially for the fast-growing cities of today’s rapidly urbanizing countries.

Our contribution is threefold. First, we develop a method to detect prime locations using big data techniques. Combining them with a new historic dataset of urban biographies, we track changes in the geography of jobs within 125 global cities over more than a century. We find that larger cities in 1900 and those which developed early transport networks had a more stable geography of prime services over the 20th century. Second, we develop an agent-based model of internal city structure that rationalizes the cross-sectional and dynamic patterns in our data. The model can, in particular, generate extreme geographic concentration patterns which we measure in a variety of empirically relevant ways. The presence of agglomeration economies and the endogenous development of transport networks deliver the relationship between city size and stability we uncover in the data. Third, we provide causal evidence for the mechanisms that drive the relationships highlighted in our model. Using detailed data on exogenous shocks from our urban biographies and a novel instrument—subway potential—we disentangle the effect of city size from that of transport networks. Both agglomeration effects and early infrastructure explain why cities that were larger in 1900 experienced less change over the following century than smaller cities did.

Analyzing changes in the within-city distribution of jobs, over a long horizon and for a large number of cities around the globe, is important but difficult. Consequently, little has been done thus far.¹ The first challenge is that, even for contemporaneous cities, we generally lack systematic data on the location of employment within cities. This problem

¹Duranton and Puga (2015, p.551) state in their recent survey: “For employment decentralization [changes in the within-city distribution of jobs], the best account is still arguably the one provided by Glaeser and Kahn (2001). It is nearly 15 years old and much of it relies on county-level information. [...] It is also disconcerting that the overwhelming majority of the little we know about employment decentralization concerns only one country, the United States, which is arguably an outlier.” See Baum-Snow et al. (2017) for recent work on employment decentralization in China.

is especially acute for prime services. Consequently, despite being more important than manufacturing for cities, they are less studied.² Although big data techniques are increasingly used to approximate population, economic activity, or city geometry at a high spatial granularity—via remote-sensed land-cover or lights data (Henderson et al., 2012; Donaldson and Storeygard, 2016; Harari, 2020)—it is not clear how to infer industry-level data, especially services, from satellite imagery capturing buildings or lights. Our new approach shows how globally available big data on prime service establishment locations, in conjunction with fine-grained employment data for a selected number of cities, suffice to sensibly approximate the spatial distribution of prime service employment and to detect where the prime locations are in global cities. We provide a battery of over-identification tests to show that our prime locations are the nuclei of urban distance gradients such as commercial rent, building height, grade-A office stock held by real-estate investment trusts, co-working spaces, social media activity, and iconic workplace amenities like Starbucks franchises. Our approach performs well and—being quite general—could be extended to approximate the within-city distribution of other industries for which detailed data are available in some cities, whereas big data are available in others.

Turning to the study of long-run changes, a second challenge is that we have little information on the historic locations of prime services in cities. To overcome this problem, we make use of historic data from our urban biographies. In particular, we have collected geo-referenced information on all cities' historic foundation places, the places of their first political institution, as well as city hall locations in 1900 and 2000. We show that 1900 city halls are the nuclei of historic urban distance gradients linked to the location of prime services. In particular, data on the construction of high buildings around that period, the historic locations of bank and insurance offices, and the spatial distribution of prime service workers from the 1880 US Census all suggest that historic prime service employment declined rapidly as one moved away from the 1900 city halls.

Using our measures of urban structure and changes therein, we then provide an exploratory analysis of the locations of, and changes in, prime locations. We first document that favorable geography—such as proximity to running water or flat terrain—is a strong predictor of the places of initial settlements, which are themselves strong predictors of the locations of 1900 city halls. The latter are, to some degree, still good predictors of prime locations today. In other words, there is substantial path dependence in the spatial distribution of economic activity within cities.³ Second, we document that path dependence has become

²Manufacturing remains the focus of much empirical research in large fields such as labor (Acemoglu and Restrepo, 2019) and trade (Autor et al., 2013). While such research is important, it is less well suited to the context of cities. There were, e.g., more than 600k prime service jobs within 2 km of the Empire State Building in 2012, more than manufacturing jobs within 45 km around that point.

³The literature on path dependence in spatial configurations has typically focused either on between-city comparisons (Bleakley and Lin, 2012; Henderson et al., 2017; Michaels and Rauch, 2018) or within-city case studies (Hornbeck and Keniston, 2017; Siodla, 2017; Brooks and Lutz, 2019).

weaker over the 20th century and that there is systematic heterogeneity across cities. This suggests there is scope for multiple equilibria in the internal structure of cities, a well-known feature of models with either firm-level increasing returns (Fujita, 1988; Krugman, 1991) or external returns to scale (Fujita and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002; Ahlfeldt et al., 2015).⁴ The existence of multiple spatial steady states should be especially relevant for prime services since the latter hardly depend on locational fundamentals, are highly mobile, and likely benefit substantially from agglomeration effects such as knowledge spillovers and information sharing (Arzaghi and Henderson, 2008). Our analysis uncovers two correlates of the degree of path dependency: early densification of population and prime services as well as early mass transit systems.

Armed with these insights, we propose a new model that can rationalize these patterns and whose predictions can be taken to data. The key challenge is that the model: (i) must deliver extreme spatial concentration of prime services, where many empty and a few hyper-dense locations coexist; (ii) must predict path dependence in internal city structure; and (iii) must allow for multiple spatial steady states—whose realizations depend on initial conditions—and shifts between those steady states in the wake of shocks. These three properties are hard to jointly operationalize in either recent quantitative spatial models or older new economic geography models. The former are ill-equipped to deliver extreme spatial concentration and usually operate outside the parameter ranges that allow for multiple spatial steady states (Ahlfeldt et al., 2015; Redding and Rossi-Hansberg, 2017). The latter are difficult to work with quantitatively and hard to map to data (Fujita et al., 1999).

We take a new intermediate route and develop an agent-based model that blends key characteristics of quantitative spatial models and new economic geography models. We augment the model with endogenous transport network formation subject to increasing returns to scale. In our model, indivisible workers and firms—both of which differ along productivity—match to produce output. Assortative matching generates a fat-tailed distribution of firm sizes so that part of the extreme concentration of employment into a few locations is driven by the concentration of employment into large firms (Ellison and Glaeser, 1997; Gabaix, 2011). Firms choose locations optimally and may benefit from agglomeration economies that can be modeled in a flexible way. The model is thus closer in spirit to new economic geography models than to quantitative spatial models where firms' boundaries and location choices arguably play no role.

⁴A strand of research shows that even temporary shocks as big as nuclear bombings do not tend to have permanent effects on the spatial distribution of economic activity across cities in a system of cities (Davis and Weinstein, 2002). At the same time, some disasters (Hornbeck and Keniston, 2017; Siodla, 2017) and temporary shocks in access to markets and institutions (Redding et al., 2010; Michaels and Rauch, 2018) appear to shift the economy between multiple equilibria. Clearly, scale matters. At a very large spatial scale, city locations and sizes appear fairly stable. At a very small scale, there are almost by definition multiple equilibria due to historical accidents. More interestingly, we show that at an intermediate spatial scale within large cities, there is scope for both path dependence, multiple equilibria, and non-trivial shifts between them triggered by large shocks. Hence, this spatial scale seems to be the one where policy matters.

Our model generates equilibria that are consistent with empirical evidence. In the absence of agglomeration effects, small differences in locational fundamentals suffice to generate a fairly monocentric city pattern that replicates historic city structures and a number of urban distance gradients (Ahlfeldt and Pietrostefani, 2019). With agglomeration effects, geographic concentration generally increases and shocks to employment centers can trigger shifts between multiple equilibria. Because the agent-based model features randomness in initial conditions and in the sequentiality of agents' decisions (matching, movement), we use a Monte Carlo approach and simulate the model several thousand times. Using a realistic geography, we construct several statistics to summarize our outcome variable—the spatial distribution of employment and changes therein—in an empirically replicable way. Using regression analysis, we then explore the conditions under which the spatial economy shifts between multiple equilibria. We find that path dependence dominates the potential for multiple equilibria if a monocentric city is large. The persistence is driven by a genuine agglomeration channel—because a larger core offers greater productivity—and a transport channel—since it is easier for larger cities to bear the fixed costs of developing a durable transport network. Our model predicts that major disasters lead to shifts away from the historic monocentric structure: cities that experience stronger shocks have more dispersed employment and prime locations farther away from the historic cores. However, path dependence dominates in cities which were already large at the time when urban economies became increasingly interactive and knowledge-spillovers were on the rise (Michaels et al., 2018). These model predictions square with the stylized facts we presented before.

Guided by our model, we then provide the first empirical test for multiple spatial equilibria within cities and differences in their degree of path dependence. The regressions we run on the simulation output of the agent-based model can be taken one-for-one to the data, which allows for a clean test of its qualitative predictions. We measure long-run changes in prime locations between 1900 and 2000—using our new methodology and the data for 125 global cities—and combine it with measures of major historic disasters from our urban biographies. For identification, we require variation in 1900 city size that is exogenous to the city-internal first-nature geography. Hence, we construct a measure of agricultural returns (caloric potential) within a city's hinterland using data from Galor and Özak (2016) and a dummy indicating colonial occupation in 1800, which captures the legacy of extractive institutions (Acemoglu et al., 2001). Exploiting these historic population shifters as instruments, we find that disasters during the 20th century lead to a more dispersed distribution of prime services today, whereas the opposite is true of historic population in 1900. An additional disaster increases the average distance between prime locations by 23%, whereas doubling the 1900 population decreases that distance by 21%. These results survive an extensive battery of robustness checks. By contrast, we do not find any effects of disasters and historic population before 1900, consistent with locational fundamentals pinning down the spatial

distribution of economic activity before the dawn of the information age.

Finally, we provide evidence on the mechanisms generating path dependence. We disentangle agglomeration-induced path dependence from transport-induced path dependence, both of which explain the anchoring effect of historic city size.⁵ To this end, we use and augment data on historic subway systems from [Gonzalez-Navarro and Turner \(2018\)](#). Since subways are endogenous to population size—given their large sunk and fixed costs—we construct a novel instrument, subway potential. Historically, the adoption of subways was a response to severe congestion problems that plagued emerging cities from the mid-19th century onwards. Railways carried oceans of people to rail termini at the city boundaries, who then often needed to connect between those termini. The geometry of those termini crucially determined how people had to move across the city, thus influencing the level of congestion in the city and hence the benefit of developing a mass transit system to curb that congestion (see [Daniels and Warnes, 2007](#), p.10, for London). We exploit the geography of railway destinations surrounding a global city to predict the location of rail termini and then construct an instrument for early subway adoption using the geometry of those predicted termini. Exploiting our subway potential instruments, combined with our previous population instruments, we find that the early development of mass transit systems has a large causal effect on path dependence over the 20th century. Conditional on shocks and initial size, cities that adopted early mass transit systems exhibit 37–59% higher concentration of prime services. Moreover, once we account for the transportation channel, the separately identified population elasticities decrease by 16–56%. This highlights the importance of both agglomeration- and transport-induced path dependency. Cities with early rapid transit systems experienced significantly less change in their internal geography during the 20th century, with more compact patterns of prime service employment and contemporaneous prime locations closer to their historic city cores.

The remainder of the paper is structured as follows. Section 2 summarizes our global cities data set, develops our methodology to detect prime locations, and provides stylized facts on the spatial distribution of prime services that motivate the further analyses. Section 3 develops the agent-based model for the theoretical analysis of multiple equilibria and path dependency in the spatial distribution of prime services. Section 4 takes the model predictions to the data and disentangles the mechanisms explaining why initially larger cities experience more path dependence. Section 5 concludes.

⁵Establishing the causal effects of historic population and transport networks ([Gonzalez-Navarro and Turner, 2018](#)) is a recurring challenge in urban economics and economic history research. Inconsequential units ([Redding and Turner, 2015](#)) and historic roads ([Duranton and Turner, 2012](#)) provide exogenous variation in transport connectivity between and within cities, but do not necessarily predict the development of a within-city transport network, which is itself endogenous to city population. [Heblich et al. \(2020\)](#) show that within-city transport networks can lead to increasing segregation of productive and residential land uses. Our contribution is to document how transport networks anchor the densest economic clusters and to propose a new instrumental variables strategy to separately identify historic population and transport effects.

2 Prime locations in global cities

We are interested in how an important type of economic activity—prime services—organizes into a small number of hyper-productive clusters within large cities. We first document the rising importance of these prime services and show their fractal geographic pattern: they are highly concentrated in cities generally; they are especially concentrated among the largest cities; and, using detailed micro-geographic data, we show that they are extremely concentrated into a limited number of very dense areas within cities. The latter—locations that account for a large share of prime service employment—are what we call *prime locations*. Second, we harness big data to predict the spatial configuration of prime services and prime locations in cities where detailed micro-geographic data are not available. Last, we combine our micro-geographic data with a novel data set of historic urban biographies for 125 world cities and look at the cross-sectional patterns and dynamic changes in prime locations. In doing so, we uncover a set of facts that have hitherto gone unnoticed. These facts motivate our model and causal analysis in the remainder of the paper.

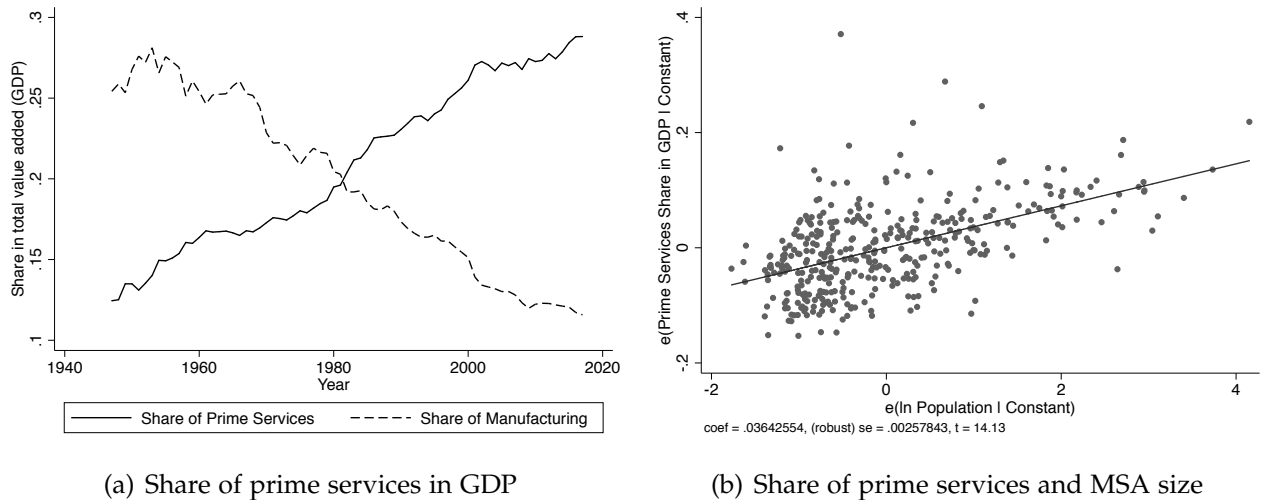
2.1 Geographic concentration of prime services

Apart from two minor exceptions, our definition of prime services encompasses all businesses in the financial, insurance, and real estate (FIRE), and professional services (PROF) sectors (see Appendix A.1 for details). These services are ‘prime’ in the sense that they are high value added, knowledge-based, and tradable; employ a large share of highly educated workers; pay high wages; and typically occupy the most up-scale office stock within cities. Today, these services make up more than a quarter of the economy. Contrasting them with manufacturing, Panel (a) in Figure 1 shows their increasing importance for US GDP over time. In terms of value added, the prime service share in the US economy grew from around 12% to over 28% from the late 1940s until today. In comparison to the almost symmetric decline of the manufacturing sector, the spatial effects of this economy-wide transformation have received relatively little attention.⁶

Data across urban and non-urban regions, between cities, and within metropolitan areas reveal the importance of prime services. Their strong urban bias operates through an extensive and an intensive margin. Even though US Metropolitan Statistical Areas (MSAs) in 2015 produced ‘only’ about 90% of the nation’s GDP, they delivered approximately 96% of prime services (Appendix A.1). In other words, prime services are almost exclusively produced in metropolitan areas. On top of this extensive margin, the semi-elasticity of 0.036 of the prime service employment share with respect to city size highlights the intensive margin of the urban bias (see panel (b) of Figure 1). Moving from the median-sized metropolitan

⁶Recent work by Eckert et al. (2020) constitutes a notable exception. The authors analyse the role of skill-intensive services for the urban bias in US economic growth since the 1980s.

Figure 1: Importance of prime services for the US economy and for large US cities



Notes: Panel (a): Calculated from the industry accounts by the [Bureau of Economic Analysis \(2018\)](#). The pattern is robust to excluding the real estate sector. Panel (b): Unit of observation is MSA. Shares are partially imputed, see [Ahlfeldt et al. \(2020b\)](#) for a description of the sources, details on the imputation, and results without imputation.

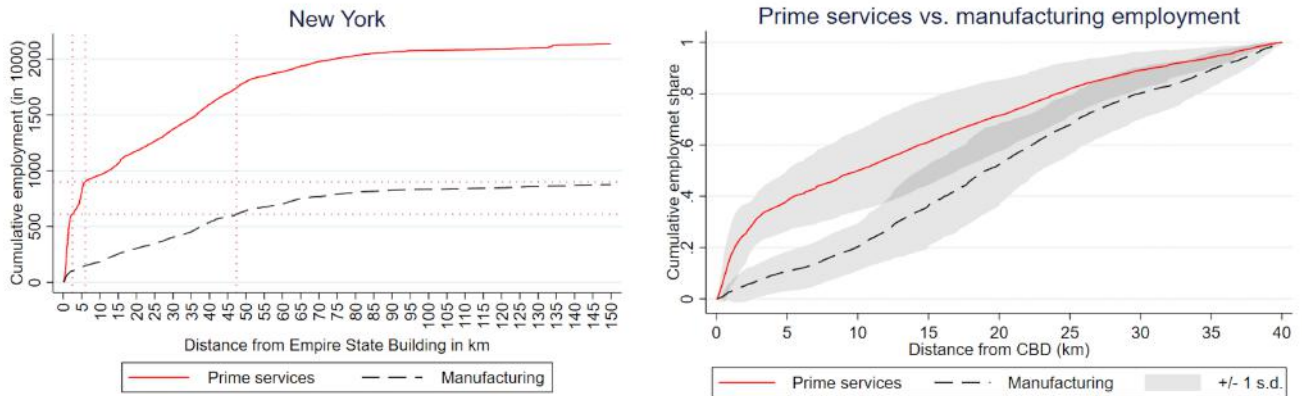
area, Champaign-Urbana, to the largest, New York-Newark-Jersey City, increases the prime service share by $[\ln(\text{pop}_{NY}) - \ln(\text{pop}_{CU})] \times \hat{\beta} \approx 4.44 \times 0.036 = 16.28$ percentage points—a change roughly equivalent to moving the US economy from 1950 to 2015 (see panel (a) of Figure 1). Great inequalities thus exist in the distribution of prime services *between* cities, with large cities being more specialized in that type of services. A priori, there are good reasons to believe that the variation in the geographic distribution of prime service employment is even larger *within* cities because various types of distance-sensitive networking externalities—agglomeration economies—matter for those services’ productivity.⁷

Panel (a) of Figure 2 shows that the within-city spatial concentration of prime services is striking. Exploiting geo-coded establishment-level employment data, the left figure reveals that prime service firms employ about 600,000 workers within a twenty minutes walk (2 km) from the Empire State Building. This number increases to close to one million if we triple the radius to one walking hour (6 km). The fact that prime service employment dominates manufacturing employment even 150 kilometers from the Empire State Building highlights its importance for a global city such as New York. More importantly, the right figure of panel (a) generalizes this insight by adding another five Canadian and US cities—Toronto, Montreal, Vancouver, Boston, and Philadelphia—for which we have geo-coded establishment-level data. The concentration observed in New York is no exception: the average share of prime service employment within 5 km from the central business district (CBD) in these six cities, computed with reference to the total within a 50 km radius, is

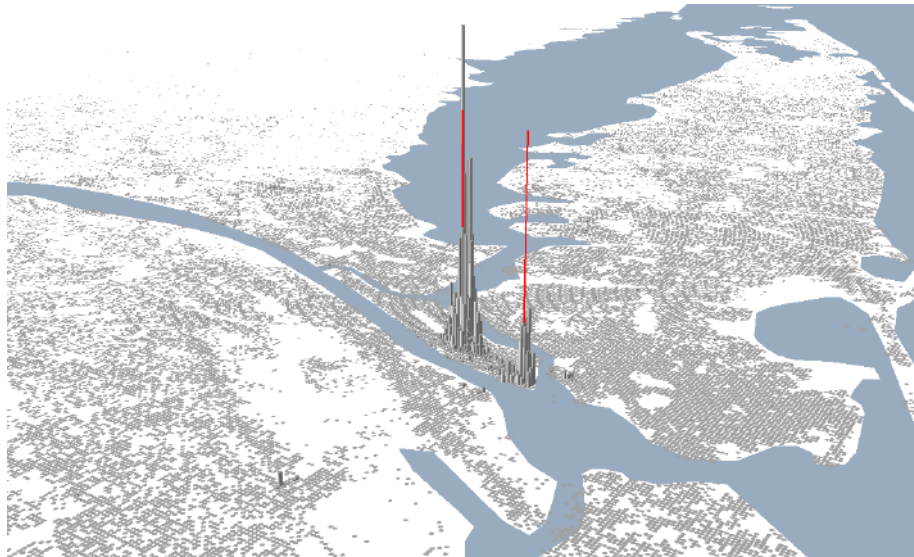
⁷See, e.g., [Arzaghi and Henderson \(2008\)](#), who study the clustering of a small subset of our prime service sector, advertising agencies, in Manhattan; and [Lui et al. \(2018\)](#), who provide evidence for within-building agglomeration economies for a subset of prime services (including law offices, advertising offices, and insurance carriers) in large US cities.

almost 40%, exceeding the respective share for manufacturing by a factor of four.

Figure 2: Observed geographic concentration of prime services in cities



(a) Distribution of prime services within cities by distance from the CBD



(b) Geographic concentration of prime service employment in NY

Notes: In Panel (a), the unit of analysis is city \times 0.1 km distance-from-CBD bins. In the right panel, we illustrate the means and standard deviations across six cities whose CBDs are defined as follows: Prudential Centre (Boston), KPMG Tower (Montreal), Empire State Building (NYC), Liberty Bell (Philadelphia), CN Tower (Toronto), and MNP Tower (Vancouver). In Panel (b), bars are proportionate to prime services employment within 250 \times 250m grid cells (white cells have zero prime service employment). Red needles illustrate global prime locations identified from prime points using our algorithm. The underlying establishment-level data are proprietary data from the National Establishment Time Series (NETS) for the US and Scott's for Canada. See the supplementary data appendix for more details.

Panel (b) of Figure 2 depicts the distribution of prime service employment in New York by 250 \times 250 meters grid cells. What is striking is the extreme concentration of those services into basically two locations—Midtown and the Financial district—whereas many other locations are completely devoid of such employment. This finding already suggests that replicating these empirical patterns theoretically requires a model that can generate extreme geographic concentration where many locations end up without any employment at all.

2.2 Predicting the location of prime services in cities

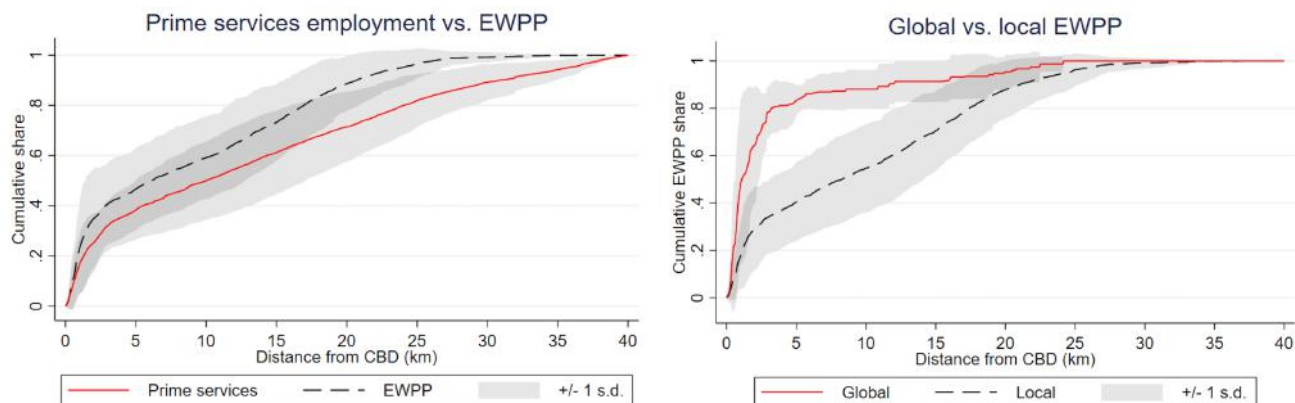
Ideally, we would rely on highly disaggregated micro-geographic data, such as those underlying Figure 2, to detect clusters of prime services. However, such data are only available for a very limited number of cities, thus restricting substantially the global scope of the analysis. Following Henderson et al. (2012), the problem of approximating the sub-national distribution of economic activity has attracted more attention. Big data techniques have, in particular, become increasingly popular in economic research as a means of approximating economic data at a high spatial or temporal granularity, e.g. via remote-sensed land-cover or lights data. Even though increasingly better resolutions open new opportunities in this field (Donaldson and Storeygard, 2016), it is nearly impossible to infer industry-level data, especially for services, from satellite imagery capturing buildings or lights. We hence propose an alternative approach and show how globally available big data on prime service establishment locations, scraped from external open sources, in conjunction with fine-grained statistical employment data for a selected number of cities suffices to sensibly approximate the spatial distribution of prime service employment for areas as small as US zip codes. Our approach complements earlier attempts to measure local economic activity using satellite imagery and extends it by allowing for a spatially fine-grained measurement of the distribution of prime services.

To construct our proxies for prime services, we combine micro-geographic data on about 457k prime service establishments for six North American cities with a dataset of about 100k prime service establishments that we scraped using the Google Places API for 125 world cities that we use in the subsequent analysis. These 100k scraped establishments—which we henceforth refer to as *prime points* (PP)—include the following categories: accounting firms, consultancies, insurance, investment banks, and law firms. We augment this dataset by scraping establishment data from the respective global leaders’ websites in these sectors (such as the top-4 accounting firms, the top-5 consultancies, and the top-10 global law firms) and locate the establishments of operating central banks and stock exchanges. These prime points serve to detect clusters of prime service employment in our sample of cities.

Unfortunately, the scraped PPs do not have any information on employment. We thus use the six cities for which we have both micro-geographic prime service establishment data—including employment—and scraped prime points to estimate employment weights for the PPs. Details of the estimation procedure and information on the estimated weights are given in Appendix A.2. In estimating those weights, we distinguish between companies with a global outreach and their more localized counterparts. One key feature of the most productive locations is that they host international firms that operate in many such places and build global networks (e.g., Sassen, 2002; Taylor, 2005). The spatial concentration of these ‘global players’ should thus particularly well reveal the places where most of the value in prime services is generated. Having obtained estimates, we finally assign the category-

specific employment weights to our prime points to obtain *employment-weighted prime points* (henceforth, EWPPs). These EWPPs will help us to predict the distribution of prime service employment in our 125 global cities.

Figure 3: Distribution of prime service employment and EWPPs in cities



Notes: Unit of analysis is city \times 0.1 km distance-from-CBD bins. We illustrate the means and standard deviations across six cities whose CBDs are defined as follows: Prudential Centre (Boston), KPMG Tower (Montreal), Empire State Building (NYC), Liberty Bell (Philadelphia), CN Tower (Toronto), and MNP tower (Vancouver). Employment-weighted prime points (EWPPs) are generated by assigning employment weights to prime points (PPs) queried from the Google Places API and scraped from company websites. Appendix A.2 documents the precise estimation procedure. Employment weights are estimated in city-specific regressions of actual prime service employment (from NETS and Scott's) observed for the six cities above against various types of prime points such as accounting firms, consultancies, insurance companies, investment banks, law firms, distinguishing between global (leading companies operating at a global scale) and local establishments, as well as central banks and stock exchanges. The unit of observation in these ancillary regressions are 30,000 disks with a 750-meter radius (about a square mile) drawn around random points. Each PP category receives the median estimated employment weight across the distribution of cities.

Figure 3 depicts the distributions of prime service employment and of EWPPs for our six cities in the same way as in panel (a) of Figure 2. In the left panel, we show that the within-city distribution of EWPPs tracks the actual distribution of prime service employment closely. Over-identification tests confirm the out-of-sample predictive power of our prime points and assigned employment weights (see Appendix A.2 for details, especially Table A 4). Importantly for the choice of a world-wide city sample explained in the next section, our EWPPs predict well the intra-city location of grade-A office buildings held by real estate investment trusts (SNL-S&P investments) and of Starbucks franchises. The former typically comprise upscale office buildings that are favored by prime service firms, especially the global players; whereas the latter arguably are an iconic workplace amenity and places where workers in knowledge-based tradable services meet and interact.⁸ Our EWPPs alone also explain more than one third of the variation in prime service employment at the level of ZIP Code Tabulation Areas within and between 39 US cities that are not used in the estimation of our employment weights. The median slope of city-specific regressions of actual prime service employment against EWPPs across these 39 cities is close to one. Last, the right panel of Figure 3 compares the concentration of companies with the largest global

⁸At the level of randomly drawn 750-meter radii disks (about one square mile), EWPPs in the six cities for which we have detailed micro-geographic data explain more than 70% of the variation in the distributions of SNL-S&P investments and of Starbucks franchises. See Table A 4 in Appendix A.2 for details.

reach with that of all other prime service companies. In line with the predictions of the global cities literature, the concentration of global EWPPs is significantly stronger, revealing that for prime services with a global reach, CBDs have a particularly strong gravity.

To summarize, our approach using employment-weighted prime points—constructed from scraped data that are available for many cities—allows us to reliably predict within-city employment patterns of prime services. The main advantage of this big data approach is that it allows us to move beyond a handful of North American cities and to investigate the geographic concentration of prime services even when no spatially disaggregated data are available. This in turn mitigates concerns about the external validity of the patterns that we dissect in the remainder of the paper.

2.3 The Global Cities dataset: Big data meets urban biographies

With a method to predict the within-city location of prime service employment in hand, we now extend our analysis to a global sample of cities. Two major challenges arise when determining and operating with such a sample: (i) which cities to choose; and (ii) how to delineate them. We chose a data-driven approach to confront both of them.

One defining characteristic of the type of cities we are interested in is their ability to attract international capital and to host large international franchises. We have explained in Section 2.2 that the locations of our EWPPs are highly correlated with those of grade-A office buildings (SNL-S&P investments) and of iconic workplace amenities (Starbucks). For both of these, we have access to an exhaustive world-wide address-level dataset of investments and franchises.⁹ We first aggregate administrative cities, as recorded in the SNL-S&P Global database, to functional cities, by selecting core administrative cities that dominate surrounding administrative cities within 30 kilometers in terms of the number of investments. We then select cities which by 2015 had at least 25 recorded investments into commercial buildings or at least 25 Starbucks franchises. After dropping a handful of cities with populations below 100k (e.g., Princeton, US) or where historic data were not traceable (e.g., Fukuoka, Japan), our resulting Global cities dataset comprises 125 cities. Our lack of data for South Asian cities and sparse African coverage aside, Figure 4 highlights that the sample is truly global. The spatial distribution of these cities follows the world map of 123 global cities in a recent report by the Global Cities Initiative very closely (Trujillo and Parilla, 2016) and covers 87% of the GDP produced in a recent top-100 cities list (PricewaterhouseCoopers UK, 2009).

Given that administrations draw city boundaries differently across the globe, it is important to define a universal rule for determining the extent of the cities. The functional cities we work with are rectangles consisting of 250×250 meter grid cells, centered on the median coordinates of prime investments of the core city. We endogenously determine the

⁹We rely on two data sources: (i) the SNL-S&P Global database, a global proprietary real estate research dataset covering grade-A office stock; and (ii) a scraped list of all Starbucks coffee shops worldwide.

Figure 4: The sample of 125 global cities



Notes: To enter the sample, a city needs to have at least 25 prime investments in grade-A office stock by real estate investment trusts in the SNL-S&P database.

side length so that we cover all prime investments within 30 kilometers of the core city, plus a five-kilometers buffer. This procedure results in endogenous city sizes ranging from 272 square kilometers for Basel to 3,875 square kilometers for Houston, which is known for its many edge cities (Lang, 2003, p. 65). We provide corresponding maps for each city in the Global Cities dataset Appendix (Ahlfeldt et al., 2020b).¹⁰

Table 1 documents the types of data we collect for our Global Cities dataset, where * symbols mark variables created from scratch, and † symbols characterize significant extensions and improvements over existing data.¹¹ Column 1 summarizes the micro-data that we employ for the prediction of current prime locations, including the core data from Google and company websites, the training data for the employment weighting, and the validation data. Column 2 summarizes geographic control variables that we employ when analyzing prime locations. Beyond first-nature geography, our analysis will also feature second-nature geography and path-dependency. The third column summarizes the variables that we collect on these dimensions. Based on many hundreds of different sources, we coded the spatial, disaster, and population history of these 125 cities. Typical starting points for creating such variables were an “urban biography” or a “historical dictionary”, which document the history of a given city. We provide a detailed account of the sources for each data point

¹⁰Recent data-intensive alternative approaches include Rozenfeld et al. (2011, delimiting US cities using a bottom-up procedure that clusters populated areas obtained from high-resolution data) and de Bellefon et al. (2019, employing building heights in the French context).

¹¹These are improvement either through calculating a new measure (e.g. deriving a market potential measure from existing global population data by (Buringh and Centre for Global Economic History, 2018; Bosker et al., 2013)) or substantial research based on urban biographies (e.g. extending and improving the data on the opening of rapid transit by Gonzalez-Navarro and Turner (2018)).

Table 1: Data summary

Disaggregated data, address- or grid-level	Urban biographies
Economic data	Physical geography
Core data	Grid level
Central banks *	Land cover
Consultancy firms *	Elevation
Co-working spaces *	Water cover
Insurance firms *	
Investment banks *	City level (built from grid)
Law firms *	Developable land †
Stock exchanges *	Irregular shape index †
	Fragmentation index †
	Caloric potential of hinterland
Employment weighting	City level
Scott’s Business Directories (P)	Colonial occupation in 1800 *
NETS National Establishment	Government type (1800-2000) *
Time Series (P)	Manmade disasters *
	Natural disasters *
Validation data	Population (1800-2000) *
Geotagged photos	Rapid transit openings †
Geotagged Twitter tweets	Subway potential (1900) †
Co-working spaces *	Market potential (1900) †
Grade-A office buildings - SNL (P)	Validation data
Starbucks establishments	1880 US census (10% sample) †
Emporis tall buildings (P)	Member list of International
US County Business Patterns	Chamber of Commerce (1922) *
	NPS historic buildings †

Notes: All economic data are at least at the address-level (many of them are geocoded at the rooftop level). * indicates primary data in the sense that we produced them from scratch. † indicates data for which we could rely on previous work and data, but that involved either substantial additional archival research (e.g., rapid transit openings) or substantial additional own calculations (e.g., developable land measure). (P) marks proprietary data. No sign means that the data were simply matched to our city dataset, such as caloric potential from Galor and Özak (2016). All sources are documented in the accompanying Global Cities dataset Appendix (Ahlfeldt et al., 2020b).

alongside with a population graph, a map showing the location of prime services, and a disaster history in the separate Global Cities dataset appendix (Ahlfeldt et al., 2020b). Finally, Table 1 also highlights that the Global Cities dataset comes in two different forms: a raster dataset, at a 1.5×1.5 km resolution, combining address-level and grid-level data; and a city-level dataset, adding information at the city level.

2.4 Detecting contemporary prime locations

In our earlier discussion of New York and a handful of other North-American cities, we assumed the location of the central business district based (CBD) on what could be considered common sense (e.g., the Empire State Building for New York). Such approach is neither satisfactory nor scalable to our global dataset. What constitutes the CBD is subjective. ‘Common’ sense in determining business districts will inevitably vary among city planners across our 125 cities, leaving room for arbitrary choices. To complicate matters further, contemporary cities are often polycentric, featuring subcenters (McMillen, 2001) and edge cities (Henderson and Mitra, 1996). When “there is no there there”, it becomes a priori difficult to pin down a CBD.

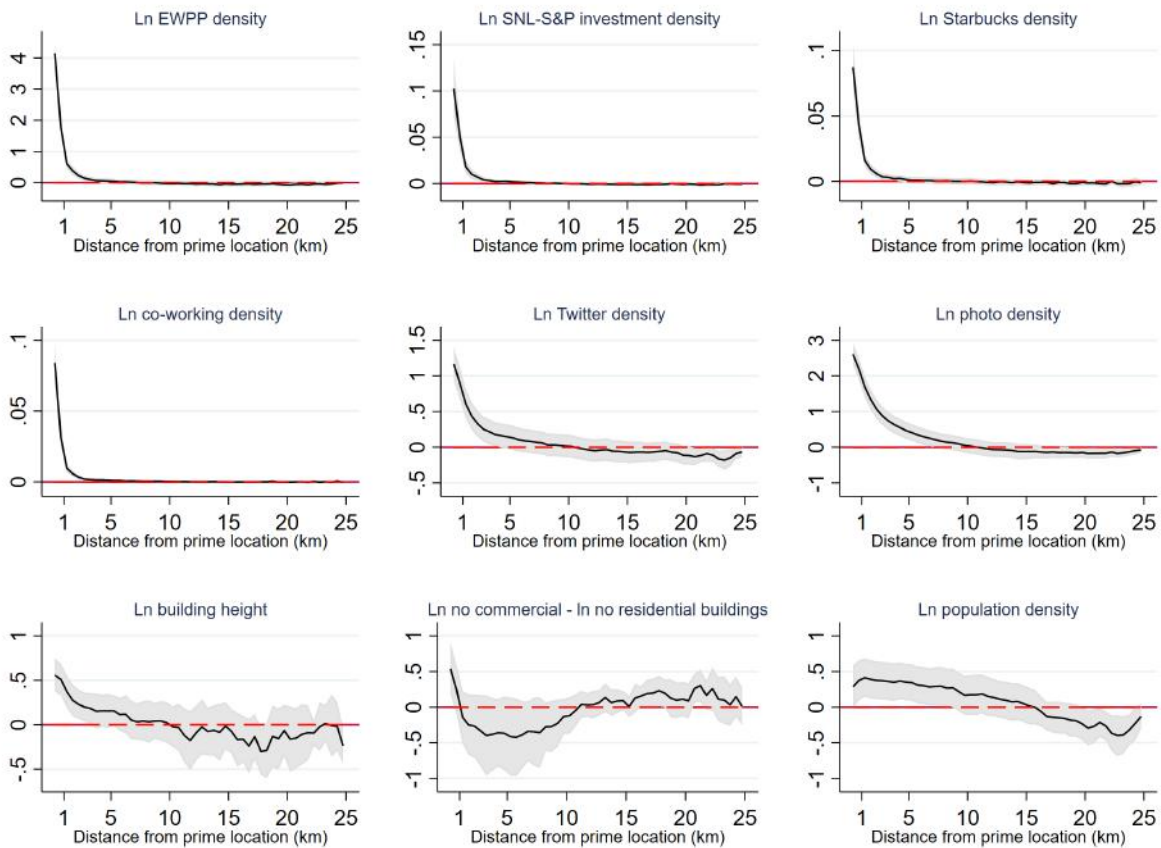
For a more systematic identification of prime locations in our dataset of 125 cities, we use our employment-weighted prime points and an algorithmic approach in the tradition of

(point-pattern based) cluster analysis (e.g., [Besag and Newell, 1991](#); [Ripley, 2005](#)). Intuitively, we detect prime locations as areas with an ‘abnormally high’ density of EWPPs compared to the expected density at any random developable point in the city. The algorithm works as follows. First, for each city we generate a distribution of 100k points uniformly drawn at random within developable cells of our city grid. Second, we loop over those 100k random points and compute the sum of EWPPs within a radius of 250 meters.¹² We also keep track of the distances of each random point from all EWPPs within that 250m radius. This yields a baseline distribution that tells us how likely any random point in the city is exposed to EWPPs. Third, we compute those same measures using the observed locations of our EWPPs instead of random points. Comparing the exposure to EWPPs of observed and random locations, we then retain only those EWPP locations in the 99th percentile of the baseline distribution. Those are locations with an abnormally high concentration of prime service employment—as measured by our EWPPs—compared to random points in the city. We refer to those points as *focal points*. Fourth, we trim away focal points in small clusters or in less dense parts of the point patterns (i.e., towards the border of the clusters) by keeping only the 25% with an above-median EWPP weight within 250m and a below-median EWPP average distance within that same radius. This ensures that focal points are in sizable and dense clusters. From the remaining focal points, we finally select those that dominate all other focal points within a catchment area of 750 meters in terms of prime point density. Prime locations are the resulting mutually exclusive catchment areas, each of which covers approximately one square mile, about the size of the City of London. With this approach, we identify 442 prime locations, which corresponds to 3.5 prime locations per global city, on average. In line with our discussion of global prime services, we replicate the procedure with the additional restriction that focal points must be within 750 meters of a global PP, which results in 240 *global prime locations*.

Figure 5 shows that prime locations are nodes of various urban density gradients. The first row shows gradients of EWPPs (which, by construction, declines steeply as we move away from PLs), SNL-S&P investments, and Starbucks franchises. We already discussed the latter two variables when justifying our choice of city sample, and Figure 5 reassures us that this generalizes more broadly to our global sample. The left panel of the second row depicts the gradient of co-working spaces—shared office spaces where workers in knowledge-based industries can interact. The gradient again falls off very quickly as we move away from prime locations. This is reassuring on two fronts. First, co-working spaces are a relatively new phenomenon and they are arguably less prone to either tourism effects (Starbucks) or the legacy of long-term investments (SNL). Second, co-working spaces are more directly

¹²In generating the 100k draws, we discard all draws that fall into undevelopable grid cells (see [Ahlfeldt et al. \(2020b\)](#) for a definition). For some draws, a 250 meters radius disk intersects with nearby undevelopable cells. In that case, we adjust our EWPP counts upward by considering the average share of developable area in the disk. When ‘counting’ prime points, we sum their weights as estimated in Section 2.2; and when computing average distances, those distances are weighted by those same weights.

Figure 5: Prime location distance gradients



Notes: The gradients are averaged across up to 125 global cities, depending on data availability. Underlying each panel is a regression of an outcome measure against 500m-distance-bin effects and city effects at the city-grid cell level. Solid black lines are the distance-bin point estimates, gray-shaded areas are the respective 95% confidence intervals. For a description of the underlying data, see the Global Cities dataset Appendix (Ahlfeldt et al., 2020b).

linked to the production of knowledge-based services. They are—by design—places that facilitate ‘social interactions for production’. That measures of social interactions—geotagged social media activity (Twitter tweets, photo counts)—also display steep distance gradients originating at our prime locations is shown by the remaining panels in the second row. Finally, the third row of Figure 5 reveals that our prime locations are also nodes of height, office space, and population gradients. The height of buildings decreases as we move away from our prime locations, as does the ratio of commercial-to-residential buildings and population density.¹³ The latter gradients show that we foremost detect dense *business* locations and not just urban density overall. The population density gradient indeed displays a different curvature from the others: it increases initially and flattens out more slowly than the other gradients.

¹³We further verified that the height gradient is steeper for recent constructions so that durable building stock alone cannot explain the density gradient. Additionally, Appendix A.3 shows that the predicted prime locations pick up concentration measured from industry-specific establishment-level data for six North American cities.

Figure 6 illustrates the inputs (employment-weighted prime points) and the outputs (global and non-global prime locations) of our cluster algorithm for the cities of Los Angeles and New York, respectively. It also provides additional validation. Within each of the cities, it becomes apparent that our algorithm identifies the most important clusters of employment-weighted prime points relative to the distribution of validation data such as Starbucks, SNL buildings, and co-working spaces. Across cities, the pattern of prime locations reproduces a well-known dichotomy: LA, hosting many prime locations afar from each other, is “beyond polycentricity” (Gordon and Richardson, 1996, as cited in [Anas et al. 1998](#)), whereas NY remains centered around two cores in Manhattan ([Glaeser, 2011](#)).

Figure 6: Prime locations in Los Angeles (left panel) and New York (right panel).



Notes: This figure illustrates the output of our cluster algorithm that identifies focal points as the most clustered prime points and prime locations as the centers of clustered focal points. Own illustration based on OpenStreetMap background map. Complete city profiles, including similar maps for all 125 cities, are available in the Global Cities dataset appendix ([Ahlfeldt et al., 2020b](#)).

Such heterogeneity in the geography of prime services is striking. NY and LA are both large metropolitan areas in the same country that share a similar institutional framework. Moreover, the dispersion in LA is a recent phenomenon—both cities were monocentric in the mid- to late 19th century. The early business district in LA is close to what is now the Civic Center area in Downtown LA ([Roseman et al., 2004](#), pp. 11 & 35). In New York, the historic central business district began at City Hall and encompassed Broadway and the Bowery according to historians and econometric evidence ([Atack and Margo, 1998](#), p. 159). In both

cities, the historic centers of prime service employment were thus close to the foundation places and 1900 city halls. Figure 6 thus nests an illustration of the LA's and NY's diverging geography of prime service employment over the 20th century. In LA, the *central point*—intuitively denoting the location with the greatest accessibility to prime points within 5 km (see Appendix A.4)—has moved halfway from Downtown towards Santa Monica, while in NY it remains close to the historic center of Manhattan. Which differences between the two cities can explain this shift?

Differences in their first-nature geographies may partly explain the divergence. For example, the Hudson and East rivers constrain Manhattan's urban form, whereas LA's historic core does not face similar constraints. The resulting differences in urban compactness may lead to differences in accessibility and have effects on agglomeration or welfare (Harari, 2020). Moreover, Lower Manhattan, unlike Downtown LA, is a natural harbour. This locational fundamental may explain why the prime locations in NY are so close to their initial settlement, whereas they moved seawards (as measured by the central point) in LA.

Similarly, the development of NY's and LA's second-nature geography differs substantially in at least two respects. First, NY was settled more than 150 years before LA, it was much larger, and much denser earlier. By 1900, it hosted 30 times the population of LA and about 5 times as many people per hectare. (Advisory Commission on Intergovernmental Relations, 1977). The higher density likely generated greater agglomeration economies, increasing the centripetal force of the initial prime location and the hurdle for new prime locations to emerge outside of Manhattan. Second, "transportation technologies shape cities" as Glaeser (2011, p.141) points out with regard to NY. Midtown hosted the Pennsylvania and Grand Central stations, connecting New York with the rest of the country. These stations also became the center of a massive rapid transport system starting to develop in 1868 as an elevated metro. The transit system, which was (and still is) largely centered on Manhattan, reinforced the locational advantage of the historic centers, thus anchoring them. LA did not develop such an early hub and spoke system, possibly because its smaller size in 1900 made it unprofitable, and thus organized around car-based transportation in the 20th century.

In sum, our algorithm reliably detects prime locations—centers of modern prime service activity. The comparison of New York and Los Angeles highlights that the contemporary spatial configuration of prime locations ranges from concentrated to very dispersed. In the case of LA and NY, both large cities nowadays, first nature geography and the history of second nature attributes, early population agglomeration and rapid urban transport in particular, seem likely candidates to explain the divergence. Can this insight be generalized?

2.5 Historical prime locations

To make more generalizable statements about the role of first- and second-nature geography, we need to move to our 125 city sample. This poses two challenges. First, we have to

determine a common date across the sample for which we gather data on historic prime service employment centers (the historic ‘prime locations’)—ideally just before second nature geography started to matter. Second, we have to locate these centers in the absence of data sources that match the extent of our scraped prime point data for today.

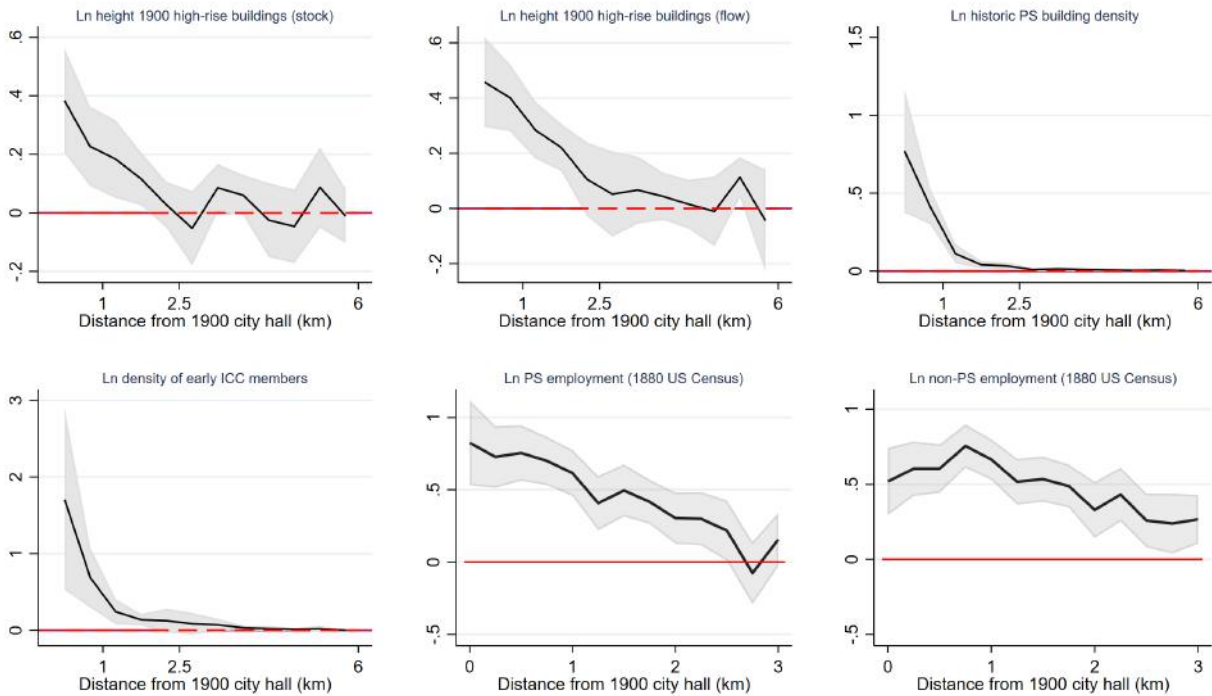
Canonical models of the economic geography literature consider the industrial era as the starting point for second nature attributes to matter. Beginning with this period, locational fundamentals become less important compared to external returns to scale that arise from input sharing, labour market pooling, and knowledge spillovers (Marshall, 1890; Krugman, 1991). From the early 19th century onwards—the periodization varies by country—industrialization facilitated the growth of cities as peasants migrated in from the countryside and firms began to exploit internal and external economies of scale in manufacturing. Demand for prime services such as banking grew as the nature of the economy changed.¹⁴ However, the spatial structure of cities remained constrained by low speeds at which people and information travelled through space. This changed towards the end of the 19th century when the electric street car and later subways, elevated and underground, facilitated urban sprawl (Muller 2017). The powerful individual transport and communication revolutions, epitomized by the rapid adoption of the automobile and telephone, followed soon after (see Figure A10 in Appendix A.6). These reductions in transport and communication costs induced urban areas to specialize according to their comparative advantage in interactive tasks (Michaels et al., 2018; Hebllich et al., 2020). Within cities, more remote places became feasible business locations (Moyer, 1977). We thus select 1900 as the end of the first nature period and argue that from there on agglomeration effects progressively took over.

Detecting historic prime locations requires a different approach as we lack the wealth of data that we use to identify contemporaneous prime locations. Urban history, however, is a good guide. The foundation place provides a good proxy for the very early city center by definition: With surrounding parts not settled or settled later, prime service activity must have taken place there. As discussed by Anas et al. (1998), the historical evidence suggests that cities, for the reasons discussed above, remained centered around one core until the end of the 19th century. This simplifies the problem since we only need to find *the* historic prime location. The 1900 city hall is a natural candidate for the historic prime location since political and economic centers could not afford to decouple due to high moving and communication costs. Indeed, Figure 7 confirms that 1900 city halls were the nuclei of various density gradients that we associate with prime services activity.

The upper-left panel shows that 1900 city halls co-located with tall buildings. This is especially true for new constructions around 1900–1910, which suggests that a 1900 city hall reveals the economic center in 1900 rather than the legacy of a more distant past (upper-center panel). The upper-right panel harnesses data from the historic places database of

¹⁴According to data constructed by Philippon (2015), the value added share of the finance and real estate sector in the United States almost tripled between 1869 and 1899.

Figure 7: 1900 city hall distance gradients



Notes: The gradients are averaged across cities. Underlying the figure are regressions of the different outcome measures on city fixed effects and 250m-distance-bin dummies (graphs containing US Census data) and 500m distance bins (all others). Outcome measure in upper-left and upper-center panels: $\ln(\text{building heights})$ in grid cell (only non-missing data included; $N = 2,174$ for 90 cities and $N = 3,457$ for 93 cities, respectively); outcome measure in upper-right (ur) and lower-left panels (ll): $\ln(\mu + \text{establishment count})$, where μ is the mean establishment count per cell ((ur) based on 36 US cities and more than 250 establishments in total (ll) based on 14 American and European cities with more than 450 establishments in total). The outcome measure in the bottom-center and bottom-right panels are $\ln((\text{number PS workers}_j + 1)/\text{area}_j)$ and $\ln((\text{number nonPS workers}_j + 1)/\text{area}_j)$. Data are for 26 cities. The solid black lines are the point estimates and the gray-shaded areas are the confidence intervals of the latter. For a more detailed description of the underlying data, see the Global Cities dataset Appendix (Ahlfeldt et al., 2020b).

the US National Park service, focusing on historic bank, insurance, and association buildings. The lower-left panel replicates this exercise with a members list from the International Chamber of Commerce from 1922, containing global address-level data of business associations and banks. These gradients for historic prime service buildings fall off very quickly with distance from the 1900 city halls. The last two panels show gradients for people working in prime services and those not working in prime services, respectively, using the 10% sample of the 1880 US census for 26 cities. The gradient for workers outside of prime service industries is not monotonic and also flatter than the corresponding gradient for prime service workers. It mimics the population gradient in Figure 5 in that increases within the first kilometer before. This comforts us in our view that the 1900 city halls pick up prime services centers at high geographic precision.

2.6 Resilience of prime locations: stylized evidence

Equipped with information on historic prime locations for 125 global cities, we are now ready to search more systematically for potential determinants of spatial change and per-

sistence in global cities. Concretely, we are seeking to substantiate or reject the role of location fundamentals, historic agglomeration, and early transport adoption suggested by our discussion of New York and Los Angeles (Figure 6). At this stage, we keep the analysis descriptive and regress binary indicators that mark the presence of contemporaneous or historic prime locations within 1.5×1.5 km grid cells against grid-cell characteristics, with no particular emphasis on causality. This exploratory analysis will guide our modelling choices in Section 3 and the investigation of causal mechanisms in Section 4.

Table 2: Determinants of prime locations

	(1) FP	(2) CH1900	(3) Global PL	(4) PL	(4a) PL	(4b) PL	(4c) PL	(4d) PL
	full sample				early skyscraper		early subway	
					yes	no	yes	no
Water in own or adjacent cell	1.634*** (0.23)	1.320*** (0.23)	0.831*** (0.18)	0.558*** (0.14)	0.422** (0.18)	0.681*** (0.22)	0.622+ (0.39)	0.549*** (0.15)
Average elevation	-12.130** (5.01)	-7.171*** (1.79)	-1.515+ (0.98)	-1.453 (1.05)	-1.706 (1.92)	-1.223 (1.29)	-5.774 (6.79)	-1.334 (1.07)
Average slope	10.332** (5.21)	2.243 (2.30)	-1.529 (1.70)	-4.207** (1.73)	-1.037 (3.42)	-6.127*** (1.91)	9.640 (7.79)	-4.916*** (1.78)
Bare ground	-1.860+ (1.18)	-1.368* (0.77)	0.397 (0.70)	0.015 (0.55)	-10.714*** (0.79)	-0.041 (0.57)	(.)	0.001 (0.55)
Foundation place in cell		5.599*** (0.24)						
Near 1900 city hall			4.098*** (0.12)	3.392*** (0.09)	3.730*** (0.13)	3.023*** (0.14)	3.634*** (0.27)	3.364*** (0.10)
Cities	125	125	125	125	70	55	14	111
<i>N</i> (cells)	111,114	111,114	111,114	111,114	62,459	48,655	12,119	98,995
McFadden's Pseudo R^2	.07	.24	.21	.12	.15	.09	.16	.12

Notes: Unit of observation are 1.5×1.5 km grid cells. Dependent variables: FP = Foundation place (in cell); Global PL = Global prime location; PL = prime location; CH1900 = 1900 city hall. Units of independent variables: water: dummy (1,0); elevation: km; slope: %, bare ground: dummy (1,0); Foundation place: dummy (1,0); Near 1900 city hall: (queen) contingency to 1900 city hall or city hall in own cell dummy (1,0). All regressions include metro fixed effects. Where no city hall in 1900 exists, we take the first city hall instead. In the rare cases where FP or CH1900 (2) lie outside our grid, we assign a 1 to the closest grid cell. Columns (4a-d) split the sample by metro-wide attributes. All models are estimated using a Poisson maximum likelihood estimator with standard errors clustered by metro. Using interactions between the 'Near 1900 city hall' variable with a dummy for the early skyscraper and early subway samples (rather than splitting the sample) yields differences of .66 (p-value=.00) and .35 (p-value: .18) respectively. For a description of the variables and the underlying data, see the Global Cities dataset Appendix (Ahlfeldt et al., 2020b). + $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In the absence of agglomeration effects, the spatial distribution of economic activity should be mainly tied to locational fundamentals, i.e. physical geography features. Column (1) of Table 2 shows that the historic foundation place tends to be in grid cells with a favorable geography, i.e., close to running water, at lower altitudes, in valleys, and on relatively flat terrain.¹⁵ Having access to nearby navigable waters, as an example, increases a grid cell's probability to host the historic foundation place by $\exp(1.634) - 1 = 412\%$. More generally, most of our locational fundamentals in column (1) are predictors of where the foundation place was located. Column (2) replicates column (1) to predict the location of the 1900 city halls. It shows that locational fundamentals continue to matter, although their effect gets weaker (e.g., water drops from 412% to 274%). Furthermore, Column (2) shows

¹⁵Soil fertility was also a very important historic determinant of where cities emerged and how large they could grow. However, it mattered little for the location of economic activity within cities. We later use soil fertility (caloric potential) as an instrument for city size.

that city structure displays substantial inertia: proximity to the historic foundation place is a very strong predictor of the 1900 CH location.

Columns (3) and (4) of Table 2 show that locational fundamentals and inertia remain important to predict where today's prime locations are: proximity to the 1900 city hall still matters. This is particularly true for the global prime locations, which are more strongly anchored to the historic city cores than prime locations generally. Note that proximity to the foundation place and the 1900 CH—both influenced by locational fundamentals—partly explain why locational fundamentals still seem to matter for prime locations. Yet, the estimated coefficients on locational fundamentals decrease in magnitude and tend to lose statistical power. The rise of second-nature geography offers a plausible explanation. As external returns to scale—which follow economic activity, i.e., are geographically mobile—become progressively more important, the spatial equilibrium distribution of employment is no longer uniquely determined: provided a sufficient mass of activity concentrates somewhere—wherever that may be—the endogenous agglomeration effects ensure that local productivity keeps economic activity put where it is. A well-known corollary is that, contrary to a first-nature world, shocks and historical accidents may have permanent effects in a second-nature world. The underlying reason is that locational fundamentals are not affected by the shock and stay put, whereas the shock-induced movement of economic activity also moves the endogenous agglomeration economies in space.¹⁶

The first four columns in Table 2 suggest that locational fundamentals are unlikely to explain why prime locations moved away from their original locations (such as in LA, where the center of gravity shifted towards the ocean). In the remaining columns, we explore the potential role of historic agglomeration and early rapid transport adoption. Measuring the size and spatial concentration of historic prime locations is not straightforward. Our strategy is to use a measure of 'revealed densification'.¹⁷ We split our sample into those cities that had a skyscraper in 1900 (4a) and those that did not (4b). The coefficient capturing path dependence (Near 1900 City Hall) shrinks by about 20% as we move from cities with early dense cores to those lacking such cores, while the pseudo R^2 drops by 40%. Dense early centers hence seem to have strongly anchored the production of prime services. To explore the effect of early rapid transport adoption, Columns (4c) and (4d) of Table 2 split the sample into those cities that were early adopters of rapid transport systems and those that were not.¹⁸ The near 1900 city hall coefficient drops by about 7.5% and the pseudo R^2 by

¹⁶In a first-nature world, once the shock is resorbed, economic activity should move back to its initial location (Davis and Weinstein, 2002). Table A 8 in Appendix A.6 shows that city structure indeed remained very stable before 1900, with 1900 city halls tightly linked to historic foundation places irrespective of the shocks that the cities experienced between 1800 and 1900.

¹⁷Early skyscraper adoption and total population are strongly correlated. Based on our global cities sample, a linear probability model yields $\beta \approx .07$ (t-statistic about 3), suggesting that the probability of having a skyscraper increases by $\ln(5/1) * .07 = 11\%$ as we move from a city with around 100,000 inhabitants to one with more than 500,00 inhabitants in 1900.

¹⁸We choose 1910 instead of 1900 to allow some time lag for construction.

about 25% as we move from the early adopters to the rest of the sample, suggesting an anchoring effect of early subway networks on contemporaneous prime locations.

To summarize, the production of prime services accounts for a large share of GDP of modern cities and is extremely concentrated within them. Prime services are produced predominantly in a few prime locations so that the dominant pattern is one of ‘cores’ and ‘peripheries’. Furthermore, while virtually all cities in 1900 were monocentric with respect to prime services, some cities changed substantially over the 20th century—becoming “beyond polycentric” like LA—whereas other cities remained very stable like NY. This suggests there is room for multiple equilibria and path dependence in city structure. Our data suggest that cities that were larger in 1900 and had denser early cores remain more strongly centered on their historic cores nowadays. This effect stems from a combination of agglomeration economies—a genuine size and density effect—and because large cities were early adopters of public transit systems. The latter anchor the historic city cores and increase the extent of path dependence. We now propose a model that can replicate these facts and whose predictions can be taken in a clean and simple way to the data.

3 An agent-based model of internal city structure

Motivated by the aforementioned facts, we develop an agent-based model (ABM) of internal city structure that blends elements of canonical quantitative urban general equilibrium models (Ahlfeldt et al., 2015) and core-periphery models of new economic geography (Krugman, 1991; Fujita et al., 1999). In line with Figure 2, our ABM generates extreme spatial concentration, with a few large clusters and zeros elsewhere. It also incorporates endogenous transport formation (as in Santamaria, 2020), multiple spatial steady states and path dependence—presumably necessary features to rationalize the observed patterns in Table 2. The crucial element to generate extreme spatial concentration is to recognize that part of that concentration arises because of the presence of indivisible large firms (e.g., Ellison and Glaeser, 1997; Gabaix, 2011).¹⁹

3.1 Building blocks of the ABM

We briefly describe the agents, the choices they make, the novel features of the model, and how an equilibrium is characterized. Due to space constraints, details, additional information on the source code, and information on the technical implementation are provided in a separate document (Ahlfeldt et al., 2020a).

¹⁹Firms are notoriously absent in quantitative spatial models that take a ‘production function approach’. Geographic patterns with zeros cannot be generated by those models, which require large exogenous productivity differences across nearby locations within a city to generate differences in the geographic concentration of production. We think it is hard to rationalize what those large location-specific exogenous productivity differences could be for the production of prime services across nearby locations within modern cities.

The basic building blocks of our ABM are locations and agents. There are two classes of agents: workers and firms. Agents are indivisible and have a unique address (a location). Workers have a unique employer, though some may remain unemployed; and firms hire mutually exclusive subsets of workers, though some may fail to hire anyone and remain inactive. Although our ABM allows for an arbitrary degree of individual heterogeneity (in terms of, e.g., housing expenditure shares, commuting decay, or locational preferences), we only parameterize those dimensions of heterogeneity that are the most directly relevant to our analysis. More details on the parameterization as well as simulation output is provided in Appendix [A.5](#).

Locations. There is a finite set of locations $\ell \in \mathcal{L}$. Each location is characterized by its land surface, geographic centroid, and a variety of parameters linked to the supply of real estate. Following quantitative spatial models, location ℓ also has a fundamental underlying productivity parameter $B_\ell > 0$.²⁰ We denote by $\mathcal{N}(\ell)$ the set of neighbors of ℓ , where a neighborhood can be defined in a flexible way.

Workers. There is a given set of workers $\omega \in \Omega$. Each worker has Cobb-Douglas preferences over housing h , and a traded consumption goods g , and spends a constant share α of his income on housing. The price of housing is r_ℓ per unit in location ℓ , whereas the price of the traded good is the same everywhere and normalized to one. Workers differ by productivity. Let θ_ω denote worker ω 's productivity.

Each worker pays commuting costs to access his workplace. If worker $\omega \in \ell$ works for firm $\varphi \in j$, the wage net of commuting costs he receives is

$$w_{\omega\varphi}^{\text{net}} = w_{\omega\varphi} \times e^{-\gamma d_{\ell j}}, \quad (1)$$

where $d_{\ell j}$ is the distance between residential location ℓ and workplace j ; $\gamma > 0$ is a commuting decay parameter; and $w_{\omega\varphi}$ is the gross wage generated by the worker-firm match (more on this below). Distance can be measured in different ways, depending on whether the city develops a public transportation network or not (we will explain this in more detail later). We use geodesic distance when moving off the transportation network and (shortest path) network distance when moving on the transportation network. Additional modeling details for workers are given in [Ahlfeldt et al. \(2020a\)](#).

Firms. Firm φ matches with a *set of* workers to produce a nationally traded prime service. We assume the price of prime services is determined outside the city and normalized to one.

²⁰Locations could also have a residential amenity parameter $A_\ell > 0$. We assume, however, that amenities differ neither across locations nor across workers and thus let $A_\ell \equiv 1$ for all ℓ and all workers ω .

Technology is such that a firm-worker match in location ℓ produces output of value

$$\mu_{\omega\varphi\ell} = B_{\ell\varphi} \times \theta_{\omega} \times \theta_{\varphi}, \quad (2)$$

where $\theta_{\omega} > 0$ and $\theta_{\varphi} > 0$ are the worker's and the firm's productivity parameters; and where $B_{\ell\varphi} = f(b_{\ell}, Y_{\ell\varphi})$ denotes a firm-specific productivity shifter in location ℓ . We assume the latter depends on a location-specific underlying fundamental productivity, B_{ℓ} , and a firm-location-specific production externality, $Y_{\ell\varphi}$. Letting $E_{\ell\setminus\varphi}$ and E_j denote the number of employees working in other firms located in ℓ and in some other location j , we assume

$$B_{\ell\varphi} = B_{\ell} \times Y_{\ell\varphi}, \quad \text{with} \quad Y_{\ell\varphi} = \frac{1}{2} \times \left[(1 + E_{\ell\setminus\varphi})^{\epsilon} + \left(1 + \sum_{j \in \mathcal{N}(\ell)} E_j\right)^{\epsilon_n} \right], \quad (3)$$

where $\mathcal{N}(\ell)$ is the set of neighbors of location ℓ over which external effects operate; and $0 \leq \epsilon_n \leq \epsilon$ is the strength of the production externality, broken down into location- and neighborhood-specific parts. Firms take $B_{\ell\varphi}$ as given. Note that expressions (2) and (3) imply that: (i) more productive workers match on average with more productive firms (e.g., [Abowd et al., 1999](#); [Bartolucci et al., 2018](#)); and (ii) more productive firms benefit more from agglomeration externalities (e.g., [Combes et al., 2012](#); [Gaubert, 2018](#)).

As explained above, we consider that a firm matches with a set of workers, i.e., there is many-to-one matching between workers and firms. To limit the size of the most productive firms, we assume there are decreasing returns in span-of-control (SOC): managing a larger number of workers generates decreasing returns beyond some point (e.g., [Rosen, 1982](#); [Eeckhout and Kircher, 2018](#)). We assume the first \bar{R}_{φ} workers generate no extra SOC costs for firm φ ; however, beyond rank \bar{R}_{φ} , hiring worker ω generates extra costs for that firm. The threshold R_{φ} varies across firms, which are hence heterogeneous along two dimensions: productivity and SOC costs. We further assume that each worker hired by firm φ requires a fixed amount ξ of office space.²¹ If firm φ hires workers $\omega \in N_{\varphi}$, it generates gross profit

$$\pi_{\varphi} = B_{\ell\varphi} \times \theta_{\varphi} \times \left(\sum_{\omega \in N_{\varphi}} \theta_{\omega} \right) - \text{SOC}(|N_{\varphi}|) - \xi \times |N_{\varphi}| \times r_{\ell} \quad (4)$$

where

$$\text{SOC}(|N_{\varphi}|) = \begin{cases} 0 & \text{if } |N_{\varphi}| \leq \bar{R}_{\varphi} \\ |N_{\varphi}|^{\gamma} - \bar{R}_{\varphi}^{\gamma} & \text{if } |N_{\varphi}| > \bar{R}_{\varphi} \end{cases} \quad (5)$$

and where $|N_{\varphi}|$ denotes the number of workers hired by the firm (the size of the firm). The first term in (4) is total revenue generated by the matches between the firm and its employees; the second term in (4) is extra cost due to SOC issues; and the third term in (4)

²¹Recent results by [Behrens et al. \(2020\)](#) using Canadian manufacturing data suggest that square footage per worker is largely constant across space within industries. Hence, at least within manufacturing, there seems to be rather limited scope for labor-land substitution. We are not aware of estimates for prime services.

is the total cost to supply the $|N_\varphi|$ workers with office space in location ℓ .

Workers and firms split the surplus generated by their match. The procedure is explained in [Ahlfeldt et al. \(2020a\)](#). Observe from (5) that there are a priori externalities between workers: when an additional worker joins a firm, he increases the SOC cost and thus reduces the firm's profit. By how much he increases that cost depends on how many workers already work for the firm. To avoid that the arrival of a new worker spills over onto the other workers' wages, we keep track of the arrival order of workers in the firm (i.e., their rank). When hiring a worker, we assume that the entire extra cost of SOC falls on that worker and does not affect the wages of workers already hired by the firm. In other words, there is no renegotiation of wages in existing matches when hiring new workers. Of course, when some workers leave the firm, the ranks of the remaining workers are adjusted—they move up in the hierarchy—and their wages increase accordingly.

Supply of real estate. We consider the simplest possible setting. Following previous work and empirical evidence (see, e.g., [Gyourko and Molloy 2015](#), for the US; and [Combes et al. 2018](#), for France), we assume there is a competitive construction sector. It combines land available in varying quantities in each location ℓ with nationally traded capital, K , to produce real estate that can be used as either housing or office space. Details are given in [Ahlfeldt et al. \(2020a\)](#).

3.2 Simulating the ABM

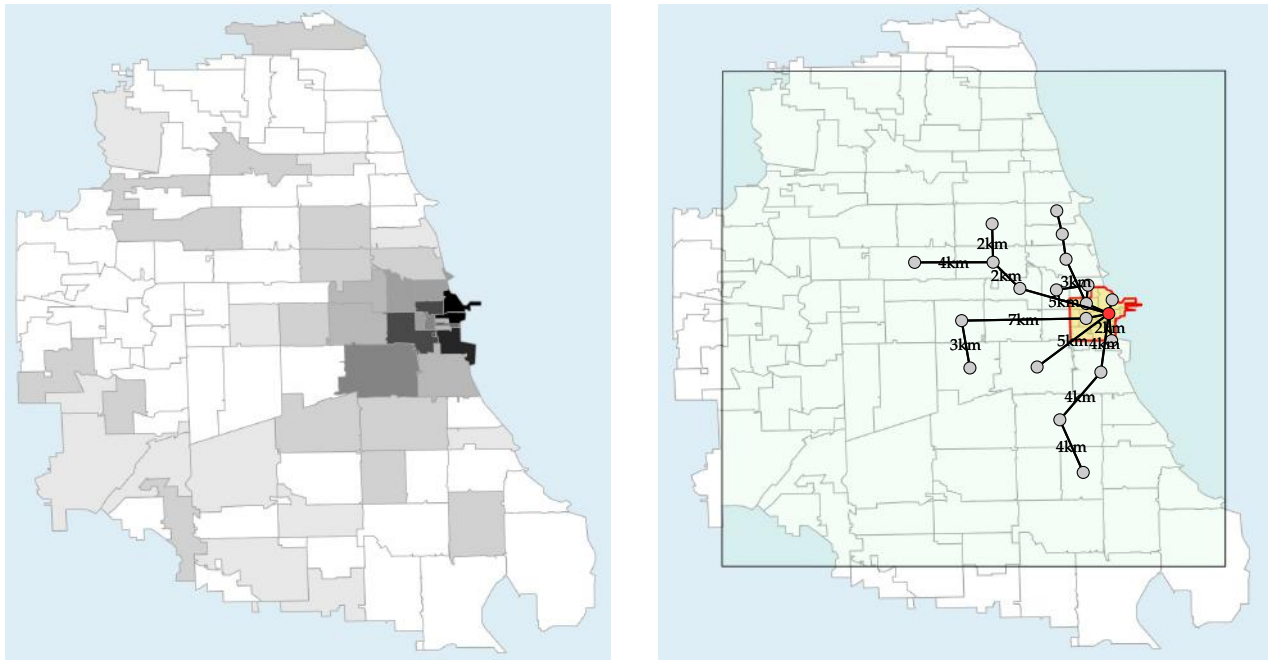
We run a series of Monte Carlo experiments—within the realistic geography of a city that consists of a set of locations in two-dimensional space—to explore how city structure shifts between multiple steady states and how city size and transportation networks shape path dependence.²² We simulate the model in the following three steps. First, we compute an initial spatial equilibrium in a world with neither production externalities nor transportation networks. In doing so, we purposely generate a 'first nature' equilibrium that resembles a monocentric city, the characteristic urban form in history. Second, we turn on agglomeration economies and move to a 'second nature' world. In that world, we allow for endogenous transportation networks and simulate shocks of random strength to the economic center of the city. We solve for a new short-run equilibrium after the shock hit. Last, we allow for the shock to dissipate and solve for a new long-run equilibrium.

We now summarize the main steps and illustrate some of the model's output. Additional model output, especially at the micro-level (worker and firm level, or location level) as well as additional validation exercises, are provided in [Appendix A.5](#).

²²For illustration purposes, we use the ZCTA (Zip Code Tabulation Area) geography of the city of Chicago—the archetype of what urban economists considered as a monocentric city in the early 20th century.

Initial equilibrium. We first simulate a spatial equilibrium in a first-nature world. We randomly assign productivity draws to workers and firms; the latter are also assigned a random SOC draw. To generate an approximately monocentric city, we use a geography with locational fundamentals in some core locations, whose strength decreases with distance from that core. We also assign residential locations to workers and locations to firms following a distance gradient from the center (see Figure A4 for an illustration). In the initial equilibrium of our first-nature world, there are no external returns to scale, i.e., $\epsilon = \epsilon_n = 0$. Workers and firms match as explained before, and firms move until no firm can profitably relocate, conditional on the worker assignment, the land rents r_ℓ (which clear the market for real estate in each location), and the locations of the other firms. More details on the procedure are given in Ahlfeldt et al. (2020a). Roughly speaking, the procedure is such that: (i) firms optimally choose locations; (ii) workers and firms can profitably rematch as firms move; and (iii) the prices for real estate adjust to clear markets. Observe that, since the initial productivity draws are random, there is variation in the initial equilibrium even when the other underlying parameters are the same. Panel (a) of Figure 8 illustrates an initial equilibrium distribution of employment.

Figure 8: Initial equilibrium and endogenously developed transport network



Notes: Darker colors in the left panel denote more employment; white ZIP Code Tabulation Areas have zero employment. In the right panel, the red node denotes the prime location of the city. The length of the links between two nodes is indicated for all links of more than 1 kilometre. The green rectangle corresponds to the empirical extent of the grid that we use for Chicago in our sample of 125 cities.

Agglomeration economies, shocks, and network formation. Starting from an initial equilibrium, we simultaneously: (i) turn on agglomeration economies ($0 < \epsilon_n < \epsilon$); (ii) allow for the endogenous development of a transportation network; and (iii) shock the center of

economic activity that emerged in the initial equilibrium of the city. To develop the network, workers' incomes and firms' profits are taxed at a uniform rate. We assume that there is a large initial sunk cost for setting up the network, and then a constant cost per station (node) and per kilometer of line built (edges). Because of the substantial sunk costs, only large cities develop a network, consistent with the empirical evidence we highlighted before. The larger the city, the larger on average the network it can develop. Panel (b) of Figure 8 illustrates the types of transportation networks that are developed. Once a network is developed, agents are free to use it to commute. They do so if it is beneficial, i.e., there is modal choice.

We also shock the prime location of the city, where the latter is assumed to be the location with the most employment in the initial equilibrium. To allow for variation in the strength of the shocks, we draw a random subset of 50-100% of firms in the prime location. The shock is modeled as a forced relocation of those firms which have to (temporarily in the short run) find a new optimal location. Firms that do not experience a shock do not move in the short run. Ahlfeldt et al. (2020a) provides details on the network formation algorithm and on shocks.

Final equilibrium. The shock dissipates in the long run and the city settles into a new equilibrium where all firms re-optimize their locations and where all workers re-optimize their assignment to firms. We refer to the final equilibrium as the 'second nature' equilibrium where agglomeration economies, shocks, and transportation networks jointly shape the spatial structure of employment centres in the city.

3.3 Simulation results

We simulate the model 10,000 times allowing for differences in: (i) city sizes; (ii) being hit by shocks and being allowed to develop a transportation network; and (iii) the magnitudes of the shocks, conditional on being hit.²³ Following the related empirical literature (e.g., Baum-Snow, 2007; Duranton et al., 2014; Harari, 2020), we summarise the spatial distribution of economic activity in each simulated equilibrium by a scalar measure at the city level.²⁴ Since there is an endless number of ways to reduce city structure to a scalar, we summarise the equilibrium distributions of each simulation run by several spatial concentration measures capturing different shades of the internal structure of the city (see Appendix A.4 for details).

²³The 10,000 simulation runs are broken down as follows: (i) 2,000 runs where we do neither allow for shocks nor for the development of a network; (ii) 2,000 runs where we allow for shocks (of varying magnitude) but do not allow for network formation; (iii) 2,000 runs where we do not allow for shocks but do allow for network formation; (iv) 2,000 runs where we have both shocks and the possibility of network formation; and finally (v) 2,000 placebo runs where we allow for shocks but do not allow for network formation and agglomeration economies in the final equilibrium.

²⁴Some of our key historic data—disasters during the 20th century and population in 1900—are not available at a detailed grid level. Working at the city level also allows us to deal with the endogeneity of historic population and public transportation, and to disregard the fact that disasters hitting one part of the city (to some cells) will affect other parts of the city via network effects.

We compute—for the initial and final equilibrium—the following six measures of effective geographic concentration: (i) the log of the average distance between EWPPs; (ii) the log of the CDF of distances between EWPPs at 1.5km; (iii) the log of the average distance between PLs; (iv) the log of the average distance of EWPPs to the 1900 city hall; (v) the distance gradient of EWPPs from the 1900 CH; and (vi) the log of the average distance of PLs to the 1900 CH.²⁵ We then run regressions to investigate the effects of city size, initial geographic concentration, shocks, and the endogenous development of a transport network on those geographic concentration measures in the final equilibrium. As these measures have an exact counterpart in the subsequent empirical analysis of our Global Cities dataset, doing so allows for a one-to-one mapping between the model-based regressions and the data. This correspondence allows us to assess in a clear way the empirical bite of the model’s key predictions.

3.3.1 Multiple equilibria vs path dependence

Our empirical approach to analyze the simulation results is as follows. Let i denote simulation run and $t = 1, 2$ denote the initial (first nature) and the final (second nature) period. Assume there exists a *pure* measure of geographic concentration of prime services, $c_i^t = f(PSempl_i^t, \mathbf{o}_i^t)$, that depends neither on city size nor on locational fundamentals. It depends, however, on PS employment, $PSempl_i^t$ that captures agglomeration economies, as well as on other potentially unobserved factors \mathbf{o}_i^t . We want to investigate how geographic concentration changes between the initial and the final stage, depending on city size and shocks. This relationship can be summarised as follows:

$$c_i^2 = \alpha(c_i^1)^\zeta (P_i^1)^\beta e^{\gamma D_i^2 + \epsilon_i}, \quad (6)$$

i.e., geographic concentration in the second nature period, c_i^2 , depends on past geographic concentration, c_i^1 , past city size, P_i^1 , and the shocks D_i^2 during the second nature period. Here, ϵ_i is a structural error term that comes from the randomness in the initial allocation of productivity across agents (i.e., even for the same P_i^1 and D_i^2 , the equilibria c_i^1 and c_i^2 may differ across runs in the ABM).

Observed measures of spatial concentration are generally not scale independent and depend on location fundamentals. Thus, we introduce a measure of *effective* geographic concentration of prime services, given by

$$C_i^t = c_i^t \times e^{a_i^t} \times (P_i^t)^{\theta^t}. \quad (7)$$

The latter includes c_i^t but also depends on locational fundamentals, a_i^t , and population, P_i^t .

²⁵As shown in Section 2.6, the 1900 city hall is located in dense areas where, especially, newly constructed buildings were high. We thus can think of the 1900 city hall as a proxy for the CBD of the monocentric city in 1900. In our model, its location is given by the (unweighted) barycenter of the firms in the city.

The latter may occur for numerous reasons: (i) large cities tend to be built more densely; (ii) large cities cover larger swaths of land and tend to develop edge-cities; and (iii) the geographic concentration measures we use can be sensitive to city size and city shape, a well-known problem in the literature. All these factors can affect the effective measure of geographic concentration of prime services, C_i^t . We have no prior on the sign and magnitude of θ^t . Substituting (7) into (6) and regrouping terms, we obtain:

$$C_i^2 = \alpha \times (C_i^1)^\zeta \times (P_i^1)^{\beta - \zeta\theta^1} \times (P_i^2)^{\theta^2} \times e^{\gamma D_i^2 + a_i^2 - \zeta a_i^1 + \epsilon_i} \quad (8)$$

which links the observed measure of geographic concentration in the second nature period, C_i^2 , to the initial measure C_i^1 , the initial population P_i^1 , the final population P_i^2 , and the shocks, D_i^2 . Observe that in the ABM $P_i^1 = P_i^2$ and $a_i^1 = a_i^2 = a_i = a$, i.e., fundamentals neither change between simulation runs nor over time. Taking logs, we can thus rewrite (8) as follows:

$$\ln C_i^2 = \ln \tilde{\alpha} + \zeta \ln C_i^1 + \tilde{\beta} \ln P_i^1 + \gamma D_i^2 + \epsilon_i, \quad (9)$$

where $\tilde{\alpha} \equiv \alpha e^{a(1-\zeta)}$ and $\tilde{\beta} \equiv \beta - \zeta\theta^1 + \theta_2$.

Table 3: City size and shock effects on spatial concentration measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Log city size	0.308*** (0.01)	0.050*** (0.00)	0.021*** (0.01)	0.085*** (0.01)	0.260*** (0.01)	0.155*** (0.01)	0.272*** (0.01)
Shock intensity (share)	-0.061*** (0.00)	-0.049*** (0.00)	-0.036*** (0.00)	-0.127*** (0.00)	-0.292*** (0.00)	-0.584*** (0.01)	-0.590*** (0.01)
Initial concentration	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,000	8,000	8,000	8,000	8,000	8,000	8,000
R^2	.376	.369	.018	.264	.569	.629	.484

Notes: Standard errors in parentheses. CH = city hall, EWPP = employment-weighted workplaces, PL = prime location. Unit of observation is the output of one run of the ABM simulation. We multiply all distance measures by -1 so that the estimated coefficients for all our dependent variables have the same sign. 50% of the runs do not allow for endogenous subway development to emulate exogenous variation. Ln city size is the log of the randomly assigned number of workers. Shock intensity is the share of employment in the initial prime location temporarily displaced by the shock. Initial CH location is determined as the location that minimizes the (pre-shock) employment-weighted distance to all workplaces. 50% of the runs exclude shocks to introduce extensive-margin variation in shock exposure. Standard errors (in parentheses) are bootstrapped in 1000 iterations. + $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3 summarises OLS regression results from (9) using our different measures of effective geographic concentration. It shows that economic activity is significantly more concentrated in the final equilibrium in larger cities as compared to smaller cities, conditional on the initial geographic concentration as captured by the lagged outcome. As further shown, shocks to the city tend to reduce the extent of geographic concentration in the final equilibrium (columns 1–3) and to shift the employment cores away from the historic cores, as proxied by the location of the 1900 city hall (columns 4–7). Hence, there is evidence for both multiple equilibria and path dependence: shocks have the potential to shift city struc-

ture between equilibria, but this is less likely to occur—and less pronounced—if the city is initially large. These findings are fully consistent with those of our descriptive grid-level analysis which reveals that locational fundamentals matter less today than they did in 1900 (see Table 2).

3.3.2 Path dependence: Agglomeration vs transport

We next explore the channels through which city size may anchor city structure in the long run. As suggested by the last two columns of Table 2, transport infrastructure may play a significant role. Table 3 did not control for transport infrastructure, although our model allows for its endogenous development. Since only large cities can afford to develop large-scale infrastructure, city size in Table 3 likely captures both externality-induced persistence and persistence due to the presence of infrastructure. To disentangle these two effects, we allow for an independent transport network effect on effective concentration and expand expression (7) as follows:

$$C_i^t = c_i^t \times e^{a_i^t} \times (P_i^t)^{\theta^t} \times e^{\delta^t M_i^t}, \quad (10)$$

where M_i^t is an indicator for simulation run i having a transport network in period t and δ^t is the respective effect. Substituting (10) into (6) and regrouping terms, we obtain:

$$C_i^2 = \alpha \times (C_i^1)^\zeta \times (P_i^1)^{\beta - \zeta \theta^1} \times (P_i^2)^{\theta^2} \times e^{\gamma D_i^2 + a_i^2 - \zeta a_i^1 + \epsilon_i} \times e^{\delta^1 M_i^1 + \zeta \delta^2 M_i^2}. \quad (11)$$

Proceeding as in the previous subsection and noting that $M_i^1 = 0$ —there is no transport network in the first stage as it can only be developed at the beginning of the second stage—we obtain:

$$\ln C_i^2 = \ln \tilde{\alpha} + \zeta \ln C_i^1 + \tilde{\beta} \ln P_i^1 + \gamma D_i^2 + \tilde{\delta} M_i^2 + \epsilon_i, \quad (12)$$

where $\tilde{\alpha} \equiv \alpha e^{a(1-\zeta)}$, $\tilde{\beta} \equiv \beta - \zeta \theta^1 + \theta_2$, and $\tilde{\delta} \equiv \zeta \delta^2$.

Table 4 summarises OLS regression results from (12). It shows that, as expected, the effects of shocks are identical to those in Table 3 since they are random and orthogonal to the other variables. Observe, however, that the extent of geographic concentration in the final equilibrium now depends positively on both city size and the presence of a transportation network. Comparing Tables 3 and 4, we see that a substantial share of the size effect goes through the transportation network.²⁶ Larger cities display more geographic concentration of economic activity and more spatial stability because of both agglomeration economies and the development of (large) transportation networks. The latter involve substantial set up costs, which explains why they are only found in large cities. As they are highly persistent and centered on the historic cores, they partly explain why city size is strongly correlated with geographic stability.

²⁶The sum of the coefficients on ‘Log city size’ and ‘Subway (0,1)’ in Table 4 is virtually equal to the coefficient on ‘Log city size’ in Table 3. This shows that we can cleanly separate the two effects in the model.

Table 4: Agglomeration vs. transport-induced path dependency

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Log city size	0.134*** (0.01)	0.015*** (0.00)	0.013* (0.01)	0.084*** (0.01)	0.125*** (0.01)	0.072*** (0.01)	0.184*** (0.01)
Network (0,1)	0.165*** (0.00)	0.033*** (0.00)	0.008*** (0.00)	0.000 (0.00)	0.127*** (0.00)	0.079*** (0.01)	0.084*** (0.01)
Shock intensity (share)	-0.059*** (0.00)	-0.048*** (0.00)	-0.036*** (0.00)	-0.127*** (0.00)	-0.290*** (0.00)	-0.583*** (0.01)	-0.589*** (0.01)
Initial concentration	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,000	8,000	8,000	8000	8,000	8,000	8,000
R^2	.615	.515	.0193	.264	.639	.638	.495

Notes: Standard errors in parentheses. CH = city hall, EWPP = employment-weighted workplaces, PL = prime location. Unit of observation is the output of one run of the ABM simulation. We multiply all distance measures by -1 so that the estimated coefficients for all our dependent variables have the same sign. 50% of the runs do not allow for endogenous subway development to emulate exogenous variation. Ln city size is the log of the randomly assigned number of workers. Shock intensity is the share of employment in the initial prime location temporarily displaced by the shock. Initial CH location is determined as the location that minimizes the (pre-shock) employment-weighted distance to all workplaces. 50% of the runs exclude shocks to introduce extensive-margin variation in shock exposure. Standard errors (in parentheses) are bootstrapped in 1000 iterations. + $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.4 Summary of theoretical predictions

The takeaway message from the ABM is as follows. With external returns, there is potential for multiple equilibria and shocks can shift city structure between steady states. Yet, if cities were initially large, then path dependence may dominate because of the persistence induced by the historic city cores. The latter effect is reinforced by the endogenous development of transport networks—which occurs mainly in large cities because of substantial set up costs—but does not entirely depend on it; there is also agglomeration-induced path dependence via external returns to scale. In a nutshell, we expect that contemporary cities that were historically larger in the first-nature era (around 1900) display more geographic concentration of economic activity today because: (i) there are agglomeration externalities; and (ii) these cities were more likely to develop large-scale transportation networks at an early period. Path dependence will dominate in cities that were already large at the time when urban economies become increasingly interactive and knowledge-spillovers were on the rise. Shocks to cities have the potential to change their internal structure. The ABM predicts that cities that experience larger shocks during the second-nature era (say 1900–2000) tend to be less geographically concentrated nowadays than cities that experienced no shocks or smaller shocks, conditional on initial size, the initial geographic concentration of economic activity, and network development.

4 Taking the model to the real world of concentration

We now provide causal evidence for the model predictions and mechanisms for our 125 global cities between 1900 and 2000. In taking the theoretical predictions to data, we are naturally concerned about unobserved confounders that are absent in our Monte Carlo simulations. We show that this challenge can be overcome by exploiting variation external to within-city characteristics and construct instrumental variables that satisfy this condition: caloric potential and colonial occupation for 1900 population and a measure of *subway potential* for the early adoption of a rapid transport system. Equipped with these instruments and our historic data on disasters, we map our numerical exercises from the theoretical section—including the outcome variables—one-for-one into the real world of concentration.

4.1 Multiple equilibria and path dependence

Following our model, we distinguish between a *historic* period, where first-nature geographic features (e.g., access to navigable waterways) primarily determine firm productivity ($t = 1$); and a *contemporary* period, where second-nature geography (e.g. knowledge spillovers) is more important ($t = 2$). All measurements are taken at the end of each period. Our reduced-form empirical specification is directly derived from (8) by taking logs and by letting $a_i^t = b^t X_i + \rho^t f_i$, where X_i and f_i capture observed and unobserved location characteristics and b^t and ρ^t allow for time-varying effects. Hence, we have:

$$\ln C_i^2 = \ln \alpha + \zeta \ln(C_i^1) + \tilde{\beta} \ln P_i^1 + \theta^2 \ln P_i^2 + \gamma D_i^2 + \tilde{b} X_i + e_i, \quad (13)$$

with $\tilde{b} = b^2 - \zeta b_1$ and $e_i = \epsilon_i + (\rho^2 - \zeta \rho^1) f_i$. Our theoretical analysis shows that major shocks to employment centers increase spatial dispersion ($\gamma < 0$) whereas historic population size anchors the distribution ($\tilde{\beta} > 0$).

To estimate (13), we construct the same geographic concentration measures, C_i^2 , as in the theoretical section for the contemporary period. Unfortunately, we do not have data to construct their historical counterparts, C_i^1 . We thus assume that, in the absence of external returns, cities display an approximately monocentric structure. This assumption implies that $C_i^1 = \bar{c} e^{\xi_i^1}$, where ξ_i^1 is an error term. An inspection of the error term, $e_i = (\rho^2 - \zeta \rho^1) f_i + \epsilon_i + \xi_i^1$, reveals that unobserved location fundamentals affecting the internal structure of cities, f_i , will lead to a biased estimate of $\tilde{\beta}$ if they are correlated with historic city size. Access to water is the case in point. As discussed in Section 2.6, it could have led to an earlier settlement and faster growth during the historic period and still attract economic activity during the contemporary period as an amenity, leading to an upward bias. However, the bias may also be downward if former amenities turned into disamenities. The reversal of the coefficient for slope in our grid-cell analysis in Section 2.6 reflects this case. It is positive for the historical foundation places, but negative for today's prime locations. A

steep slope may have been a locational advantage historically—e.g., for defending the city—but in modern times, steep slopes increase construction costs and may make it difficult for a prime location to interact with other parts of the city. In sum, we require instruments IV_i^P for historic population in 1900 P_i^1 that are orthogonal to the city-internal geography f_i to address the potential bias from unobservable within-city characteristics. Furthermore, we have to ensure that our error in capturing shocks (ϵ_i) is uncorrelated with our instruments IV_i^P and our shock measure D_i^2 , and that the latter is uncorrelated with f_i .

The real-world equivalent for D_i^2 , denoting shocks since 1900, are data from our urban biographies on disasters that caused major damage to the cities in our sample. These range from man-made disasters such as the effects of civil and international wars (e.g. the bombing of Berlin in World War II) and fires (e.g. the Great Atlanta fire of 1917) to natural disasters such as major earthquakes (e.g., the Great Kanto earthquake of 1923), hurricanes (e.g. Great Hong Kong Typhoon in 1937), and flooding (e.g., Chongqing flood of 1981). This granularity of the disaster coding will allow us to exclude potentially endogenous disasters such as fires.

What instrumental variables for population in 1900 would satisfy the requirements set out above? Fortunately, the determinants of early city growth are relatively well understood. Two of those are surely orthogonal to within-city characteristics: agricultural productivity and recent colonial occupation. First, the agricultural productivity of the hinterland shaped the growth prospects of pre- and early industrial cities (Bairoch, 1988). Analysing the introduction of the potato as a major staple across the world, Nunn and Qian (2011) indeed provide strong quantitative evidence for this mechanism at the national and city level. Going beyond the nutritional value of potatoes, we compute a measure of the expected agricultural returns within a city’s hinterland using data on *caloric potential* by Galor and Özak (2016). Like Nunn and Qian (2011), we set the radius of the hinterland to 100 kilometers. However, to exclude any impact of a city’s internal geography, we calculate the average agricultural potential within a 50–100 kilometers ring around the historic city center. A second important factor for urban growth external to cities was independence. Even though the effects of colonialism on urbanization were complex, many cities in colonial territories did face an additional obstacle relative to their counterparts in independent territories. Particularly in places with small settler communities, institutions were of extractive nature and thus uncondusive to growth (Acemoglu et al., 2001). We thus add a zero-one dummy for cities lying within *colonial territories* in 1800 as a second instrument since we expect a smaller 1900 city size for a given level of *caloric potential*, owing to the legacy of extractive institutions.

Both instruments satisfy our exclusion criterion for being external to the within-city structure in a cross-sectional dimension. The temporal dimension of our analysis makes an even stronger case for them. In a world with high transport costs—i.e., the world until the last part of the nineteenth century due to the lack of the widespread adoption of the steam engine—the caloric potential of the hinterland was more important back then than

100 years later. Likewise, the effects of colonial occupation on city sizes should have gradually faded over time as countries and cities became progressively independent. Both of these conjectures are indeed borne out in our data. While both instruments are strong predictors of city population at the beginning of the 20th century, they are not significantly correlated with population at the end of the 20th century (see Appendix A.7). This makes it less likely that they affect contemporaneous outcomes other than through path dependence.

Table 5: 1900 population and disaster effects on spatial concentration measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs ($\times -1$)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs ($\times -1$)	Distance from 1900 CH gradient ($\times -1$)	Ln dist. from EWPP to 1900 CH ($\times -1$)	Ln mean dist. from all PLs to 1900 CH ($\times -1$)	Ln mean dist. from global PLs to 1900 CH ($\times -1$)
Ln population 1900	0.213*** (0.05)	0.312*** (0.08)	0.333*** (0.07)	0.201*** (0.05)	0.255*** (0.06)	0.347*** (0.10)	0.284*** (0.11)
Disasters since 1900	-0.122*** (0.04)	-0.125** (0.05)	-0.206*** (0.06)	-0.140*** (0.04)	-0.128*** (0.04)	-0.173*** (0.06)	-0.052 (0.08)
Ln population 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. IV	CP & CT	CP & CT	CP & CT	CP & CT	CP & CT	CP & CT	CP & CT
Kleinb.-Paap F (p-val.)	0	0	0	0	0	0	0
Hansen J (p-val.)	.06	.644	.431	.748	.215	.707	.428
Observations	125	125	125	125	125	125	125

Notes: Unit of observation is cities. Columns (1) and (5) employ weighted distances. 2SLS estimates. CD = cumulative density; CH = city hall; EWPP = employment-weighted prime point; PL = prime location; CP = caloric potential; CT = colonial territories. We multiply all distance measures by -1 so that the estimated coefficients for all our dependent variables have the same sign. A unit is added to bilateral distance between PLs before taking logs so that 30 cities with only one PL obtain log values of zero. Controls include: Developed area 2000, irregular shape index, fragmentation index, share of land not developable within 5km of 1900 city hall, distance from 1900 city hall to water, share of land with steep slope. See the Global Cities dataset Appendix (Ahlfeldt et al., 2020b) for sources and data construction. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 presents estimates of specification (13) and is our empirical counterpart of Table 3. In line with our model predictions, disasters occurring after 1900 lead to more dispersed cities, whereas a larger 1900 city size anchors the spatial structure of the city, leading to more concentration around the historic core. These results lend empirical support to the prediction that in a world with agglomeration effects, major shocks can shift a spatial economy to an alternative (more dispersed) steady state, whereas historically large agglomerations increase the degree of path dependence.

In column (1) of Table 5, we use the average bilateral distance between EWPPs as an intuitive measure of spatial concentration. Doubling 1900 city size has a $2^{-0.213} - 1 = -14\%$ effect on this distance, whereas an additional disaster increases that distance by $e^{0.122} - 1 = 13\%$. In column (2), we use the cumulative density of bilateral distances between EWPPs at 750 metres to capture how much of a city's prime service activity co-locates within a very short distance, such as within a prime location. Doubling 1900 city size increases the share of EWPPs in close proximity by $2^{0.312} - 1 = 24\%$, while an additional disaster reduces that same share by 13%. In column (3), we restrict our attention solely to prime services within prime locations for which we compute the unweighted average bilateral distance. Compared to

column (1), the effects increase significantly in magnitude—doubling population decreases the distance by 21% and an additional disaster increases it by 23%. The estimated effects are sizable in light of the variation in our data. As an example, NY in 1900, with a population of 3.4 million, was about 11.5 times larger than San Francisco (299k). Since 1900, San Francisco (SF) experienced two major earthquakes that led to urban re-development, whereas NY experienced only one such event—9/11.²⁷ Jointly, the 1900 city size and disaster effects imply that, *ceteris paribus*, NY should have a more than 127% higher share of nearby EWPPs and a distance between prime locations that is 79% shorter, on average.

The remaining columns (4)–(7) of Table 5 show that the historic business district, as proxied by the city hall in 1900, exhibits a greater gravity on the distribution of prime services in cities that were historically large and experienced fewer disasters. To quantify the effects, we again compare the predicted outcomes for NY relative to SF. *Ceteris paribus*, the joint effects of historic city size and disasters imply that in NY, the rate of decay of EWPPs density in distance increases by 63 percentage points (column 4); the employment-weighted distance from PPs to the 1900 city hall is 60% shorter; the average distance from (global) prime locations to the 1900 city hall is (55%) 76% shorter. It is worth noting that the effect of disasters on the relative location of global prime locations is weaker, both in a statistical and an economic sense, than on any of the other spatial concentration measures that we employ. This suggests that the largest urban centers are the most resilient to shocks, consistent with initial conditions and size leading to path dependence and reducing the potential for multiple equilibria.

In line with our theoretical priors, we thus find substantial evidence for multiple equilibria (the disaster effect) and path dependence (the population effect). This result is robust to a battery of robustness checks, with a natural starting point being the OLS estimates (see Appendix A.8). As discussed before, the direction of an eventual bias in OLS is *a priori* unclear. While the coefficients for the disaster variable remain virtually unchanged—suggesting that these shocks are truly exogenous—those for population shrink by about a third for the OLS estimate relative to the IV estimate. This implies a downward bias in OLS, most likely owing to the presences of historic amenities that eventually turned into disamenities for contemporary cities (see our prior discussion). Appendix A.9 provides further robustness exercises. First, we show that the population effect is robust to using a single instrument (caloric potential) rather than two, even though the effects are slightly less precisely estimated. Regarding the disaster variable, we provide supplementary material to show that our results are not driven by: (a) fires, the likelihood and magnitude of which might be related to historic population density and thus not exogenous; (b) natural disasters, or (c) man-made disasters

²⁷The very destructive 1906 earthquake in San Francisco provided a “clean slate” for urban land use (Siodla, 2017). Perhaps less present in the collective memory, the Loma Prieta in 1989 led to important urban redevelopment in San Francisco such as the the deconstruction of the Embarcadero Freeway (Godfrey, 1997). To be sure, 9/11 did not destroy major parts of the city, but was a significant shock to office space, leaving 28 million square feet of office space either temporarily or permanently unusable (Haughwout, 2005).

only; and (d) recent disasters that might only cause a temporary deviation from a unique steady-state in contemporary cities, as in [Davis and Weinstein \(2002\)](#). Finally, we analyze in more detail the source of our variation. We identify the effects separately from within- and between-country variation and show that our results are not driven by North American cities only.

4.2 Path dependence: Agglomeration vs. Transport

In our model, city size-induced path dependence in internal city structure arises from two complementary forces. First, there is a genuine agglomeration effect since a more established centre in a historically larger city is more resilient to shocks. Second, durable transport networks—that are endogenous to historic city size due to substantial sunk and fixed costs—create a permanent accessibility advantage for the historic centre. To allow for an independent transport network effect on effective concentration, we make use of (11) and rewrite it as in (13) as follows:

$$\ln C_i^2 = \ln \alpha + \zeta \ln(C_i^1) + \tilde{\beta} \ln P_i^1 + \delta^1 M_i^1 + \theta^2 \ln P_i^2 + \gamma D_i^2 + \tilde{b} X_i + \tilde{\delta}^2 M_i^2 + e_i, \quad (14)$$

with $\tilde{b} = b^2 - \zeta b_1$, $\tilde{\delta}^2 = \zeta \delta^2$, and $e_i = \epsilon_i + (\rho^2 - \zeta \rho^1) f_i$.

In our empirical analysis, M_i^t is an indicator for city i having a subway system in period t and δ^t is the respective effect. In practice, concerning the network in the first-nature period, M_i^1 , we choose 1910 as the cutoff year to allow for the time lag between planning and implementation of subway systems. We expect a positive historic effect of early subway adoption, $\delta^1 > 0$. Finally, as before, we assume that cities in 1900 were approximately monocentric, i.e., $C_i^1 = \bar{c} e^{\xi_i^1}$, where ξ_i^1 is an error term.

Complementary to our discussion in the previous section, an instrumental variable IV^M for M^1 that is orthogonal to city-internal geography solves the problem of an omitted variable with respect to fundamental factors that may determine historic city size (e.g., a natural harbor that anchors the city structure). To separately identify historic population and historic subway effects, we further require that IV^M is excludable with respect to P^1 and IV^P is excludable with respect to M^1 . To find a suitable instrument for the adoption of subway systems, it is important to understand *which* problem these systems solved and *how*. Before the advent of intra-urban rapid transit, city sizes were typically limited by the ability to walk them ([Bairoch, 1988](#)), implying a maximum extent of about 50 square kilometer and a population threshold of around 475,000 (see Appendix A.6). When cities grew beyond this threshold, demand for transportation would rise. Horse carts, and later street cars, were first attempts to satisfy the demand and widen the radius of the cities. The efforts to improve urban transport culminated in the second half of the 19th century with the advent of rapid urban transport systems, either over- or underground ([Daniels and Warnes, 2007](#)). Yet, such intra-urban transportation systems were expensive and required large investments

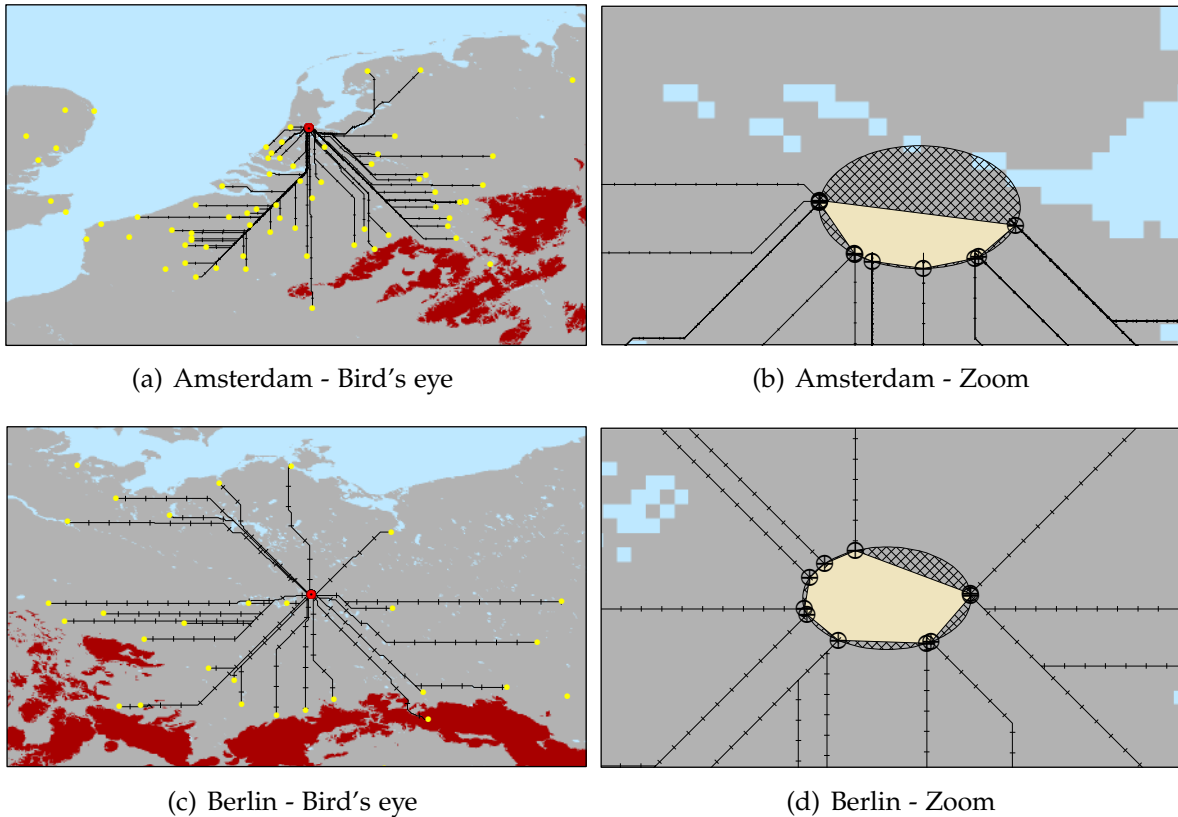
of both sunk and fixed nature. What, then, determined whether or not cities built a subway system after crossing the threshold? Besides the size of the city, the level of congestion also depended on a second factor. The number of unconnected ports of entry and exit, namely railway termini, the distance between them, and their spatial configuration were crucial for the city's congestion—and the need to solve it with rapid urban transport systems. Across the globe, early private railway companies had built their termini as close to the city as land prices and various regulations permitted. For this reason, many of them were located around the center of the city, but not at the center itself. Inevitably, the more dispersed these railway stations were, the more vehicular congestion would occur as goods and people had to move through the city—a problem that London alleviated by connecting the railway termini (Daniels and Warnes, 2007, p. 10). However, the problem and its solution were not specific to London. Indeed, a contemporary study regarding the “railway terminal problem of Chicago” highlighted that elsewhere (e.g., Boston, London, Berlin, Paris, Vienna, Hamburg) connections between the termini, in particular via circle lines, alleviated the congestion problem (City Club of Chicago, 1913, p. 84f).

These insights set the stage for constructing an instrument that predicts the early adoption of public transit but is exogenous to within-city characteristics and excludable with respect to the population instruments. In a first step, we predict the number and geography of railway termini. To do so, we draw on population data at the dawn of the railway age in 1850 to locate potential neighboring cities that could be connected with a city in our dataset (Bosker et al., 2013; Reba et al., 2016). We restrict the potential railway connections to those cities that have a great circle distance of less than 300 kilometers and a population over 15,000. Based on reasonable assumptions for crossing mountainous terrain and a high penalty for crossing water, we employ a least-cost-path algorithm to draw railway lines from the neighboring cities to our cities of interest.²⁸ We keep the neighboring cities only if their distance remains below the equivalent of 300 land kilometers. Panels (a) and (c) of Figure 9 show the corresponding ‘bird’s eye’ maps for Amsterdam and Berlin. Yellow dots mark those cities with a great circle distance below 300 kilometers, but not all are connected due to geographic constraints (such as Dover and Amsterdam, for example).

In a second step, we employ the geography of predicted railway termini to create a measure of *subway potential*. We draw a circle with radius 4km around the historic city center, which reflects the areal limits of cities before the arrival of mass transit (see above and Appendix A.6). We then compute the subway potential as the area of the convex hull of the predicted rail termini, relative to the area of the circle. Panels (b) and (d) of Figure 9 provide corresponding illustrations for Amsterdam and Berlin. Despite having crossed the population threshold to build a rapid transit system at the dawn of the 20th century, Amsterdam had a much lower probability to do so than Berlin according to our measure of subway

²⁸This is done using the ArcGIS least-cost-path algorithm. See the Global Cities dataset Appendix (Ahlfeldt et al., 2020b) for a technical exposition, the parameter choices, and a discussion of the source material.

Figure 9: Subway potential of Amsterdam and Berlin



Notes: Black lines are least-cost paths to any city with a population of at least 15,000 in 1850 (the yellow dots). Light-blue shaded areas are oceans, rivers, and lakes. Dark-brown shaded areas are mountains. The grey-shaded areas are the remaining land mass.

potential. This is corroborated by the historical record. Whereas Berlin had adopted rapid urban transport as early as 1872, the Amsterdam metro system would not open until more than 100 years later.

The subway potential that we introduce as a novel instrument for the early adoption of rapid transit systems exploits the spatial configuration of the locations of cities $j \neq i$ relative to city i . It is thus unlikely to be correlated with city-internal fundamental factors, f_i . Does it, without further controls, fulfill the second condition, i.e., is IV^M excludable with respect to historic population P_i^1 ? To rule out an effect of subway potential on historic population through a market-access channel, we condition on a historic market access measure that takes into account all global settlements with a population of at least 5,000 in 1900 and land-cover-dependent (water, plains, mountains) trade costs (following Donaldson and Hornbeck, 2016).²⁹ This allows to separately identify the genuine size effect—through agglomeration externalities—and the effect through transportation networks found in the theory.

Our instrument now satisfies all exogeneity conditions, but for it to be relevant we have

²⁹Unlike Donaldson and Hornbeck (2016), we include international market potential. The Global Cities dataset Appendix (Ahlfeldt et al., 2020b) provides details on the technical implementation and validity of the measure.

to model one more determinant. Both our theoretical model and the descriptive empirical evidence (see Appendix A.6) suggest that a minimum city size is required to economically sustain a subway system. We thus have to allow for an interaction of historic city size and subway potential whose functional form is a priori unknown. In a procedure similar to Sequeira et al. (2019), we first predict historic city size and the propensity of having a subway system using non-parametric interactions of our historic population instruments and the rail potential in two *zero-stage* regressions (see Appendix A.10 for details). These zero-stage regressions deliver two instruments for the two endogenous variables (population and subway) that are input into a conventional 2SLS approach. In keeping with expectations, there is a strong interaction between caloric potential and subway potential in the subway potential, whereas such an interaction does not exist for the population instrument. We document this and the power of the instrument in Appendix A.10.

Table 6: Agglomeration- vs. transport-induced path dependency

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.097*** (0.03)	0.207*** (0.05)	0.215*** (0.05)	0.088*** (0.03)	0.144*** (0.03)	0.291*** (0.07)	0.320*** (0.09)
Subway in 1910	0.465*** (0.14)	0.465** (0.23)	0.724*** (0.27)	0.526*** (0.14)	0.507*** (0.15)	0.572* (0.29)	0.535 ⁺ (0.34)
Disasters since 1900	-0.103*** (0.04)	-0.122** (0.06)	-0.190*** (0.06)	-0.128*** (0.04)	-0.114*** (0.04)	-0.184*** (0.06)	-0.084 (0.08)
Ln population 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subway 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market access	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. & subway IV	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage
Kleinb.-Paap F (p-val.)	0	0	0	0	0	0	0
Observations	125	125	125	125	125	125	125

Notes: Unit of observation is cities. Columns (1) and (5) employ weighted distances. 2SLS estimates. CD = cumulative density at 750 m; CH = city hall; EWPP = employment-weighted prime points; PL = prime locations. We multiply all distance measures by -1 so that the estimated coefficients for all our dependent variables have the same sign. Geographic controls include: Developed area 2000, irregular shape index, fragmentation index, share of land not developable within 5km of 1900 city hall, distance from 1900 city hall to water, share of land with steep slope. See the Global Cities dataset Appendix (Ahlfeldt et al., 2020b) for sources and data construction. Instruments for 1900 population and 1900 subway are two variables predicted in two zero-stage locally weighted regressions using population in 1900 and an indicator for subways in 1910 as dependent variables and non-parametric interactions of caloric potential, the colonial territory (in 1800) indicator, and subway potential as predictors. For details on the construction of the instrumental variables, see Appendix A.10. Robust standard errors in parentheses. + $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6 presents estimates of specification (14) and is our empirical counterpart of Table 4. It shows estimates using a zero stage that accommodates interactions between our exogenous historic population and subway shifters by means of locally weighted regressions (LWR) with narrow bandwidths that maximise the power of the first stage. Ceteris paribus, a historic subway system decreases the average distance between prime service jobs by 37% (column 1) and the average distance between PLs by 52% (column 3). The share of neighbouring EWPPs (within 750m) increases by 59% (column 2). The rate of decay of EWPPs density in distance increases by 53 percentage points (column 4). The average distance be-

tween EWPPs and PLs to the 1900 city hall decreases by 40–44% (columns 5–7), although the effect on global prime locations is borderline insignificant. Compared to the main results in Table 5, and fully consistent with our theoretical priors, the estimated historic population elasticities drop by 16–56% depending on the outcome (except for the global prime locations distance), but remain sizable.

We verify the robustness of our estimates employing alternative zero stages in which we use standard bandwidths or fully parametric interactions. Those yield similar effects, although the confidence intervals broaden. Compared to the results in Table 6, the transport effect remains significant for all outcome measures but those in columns (2) and (4) as well as the global prime locations (see Appendix A.11). Indeed, even the OLS estimates suggest a role for transport: we find positive effects for most of the concentration measures, even though these are less precisely estimated and hard to interpret given the obvious endogeneity concerns (see Appendix A.8). On balance, our interpretation of the evidence is that city-size induced path dependence in the internal organization of cities originates from agglomeration *and* durable transport networks that were endogenous to city size a century ago. Global prime locations seem to be less anchored by transport, potentially because the inertia of a densely populated center dominates.

5 Conclusion

As the growing industrial cities of the developed world did at the dawn of the 20th century, many mega cities in the rapidly urbanizing developing world suffer nowadays from congestion, pollution, and affordability problems. Hence, urban planners often wish to alter the geometry of these cities for reasons of efficiency, or equity, or both. In particular, policymakers sometimes wish to promote the emergence of new business centers to curb congestion in established cores and to revitalize economically struggling areas. This might work in theory. Indeed, a voluminous theoretical literature on agglomeration has shown that multiple spatial equilibria and path dependence are prevalent features of modern economies when there are either internal increasing returns or external agglomeration effects. It is also well understood that scale matters. At a large regional scale, history and empirical research teach us that such policies are doomed to fail. The spatial macro structure of economic activity is highly persistent and largely beyond the reach of policy—Germany’s eastern Länder or Italy’s mezzogiorno can testify to this.

We know much less at a smaller spatial scale. Our findings show that the structure of large cities is prone to multiple equilibria and path dependence, thus suggesting a potential role for policy that seeks to achieve more sustainable urban forms. Changing urban form seems possible, but is ambitious if the existing cores are well established and at the center of public transportation networks. Attempts at decentralization in big cities that have developed public transit systems—urban projects of pharaonic proportions such as ‘new New

Cairo' in Egypt—are likely to prove unsuccessful. The investments required to create new prime locations outside the established cores are likely to be too large to attract enough businesses to emulate the endogenous production spillovers that prime locations offer. Reducing the growing number of new ghost towns and new ghost business centers—such as the Yujiapu Financial District in China—that increasingly litter the landscape in urbanizing countries would certainly be important to save resources and to improve welfare. Future work should investigate whether the success of new urban developments is systematically higher in places that were initially smaller and that did not yet develop mass transit systems.

References

- Abowd, John M., Francis Kramarz, and David N. Margolis**, “High wage workers and high wage firms,” *Econometrica*, 1999, 67 (2), 251–333.
- Acemoglu, Daron and Pascual Restrepo**, “Robots and jobs: Evidence from US labor markets,” *Journal of Political Economy*, 8 2019.
- , **Simon Johnson, and James A. Robinson**, “The colonial origins of comparative development: An empirical investigation,” *American Economic Review*, 2001, 91 (5), 1369–1401.
- Advisory Commission on Intergovernmental Relations**, “Trends in metropolitan America,” Technical Report M-108, Advisory Commission on Intergovernmental Relations 1977.
- Ahlfeldt, Gabriel M. and Elisabetta Pietrostefani**, “The economic effects of density: A synthesis,” *Journal of Urban Economics*, 2019, 111, 93–107.
- , **Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf**, “The economics of density: Evidence from the Berlin Wall,” *Econometrica*, 2015, 83 (6), 2127–2189.
- , **Thilo Albers, and Kristian Behrens**, “Prime Locations: Agent-based Model Appendix,” 2020.
- , —, and —, “Prime Locations: Global Cities Data Appendix,” 2020.
- Anas, Alex, Richard Arnott, and Kenneth A. Small**, “Urban spatial structure,” *Journal of Economic Literature*, 1998, 36 (3), 1426–1464.
- Arzaghi, Mohammad and J. Vernon Henderson**, “Networking off Madison Avenue,” *The Review of Economic Studies*, 10 2008, 75 (4), 1011–1038.
- Atack, Jeremy and Robert A Margo**, ““Location, location, location!” The price gradient for vacant urban land: New York, 1835 to 1900,” *The Journal of Real Estate Finance and Economics*, 1998, 16 (2), 151–172.
- Autor, David H., David Dorn, and Gordon H. Hanson**, “The China syndrome: Local labor market effects of import competition in the United States,” *American Economic Review*, 2013, 103 (6), 2121–2168.
- Bairoch, Paul**, *Cities and economic development: From the dawn of history to the present*, University of Chicago Press, 1988.
- Bartolucci, Cristian, Francesco Devicienti, and Ignacio Monzón**, “Identifying sorting in practice,” *American Economic Journal: Applied Economics*, 2018, 10 (4), 408–438.

- Baum-Snow, Nathaniel**, “Did highways cause suburbanization?,” *The Quarterly Journal of Economics*, 2007, 122 (2), 775–805.
- , **Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang**, “Roads, railroads, and decentralization of Chinese cities,” *The Review of Economics and Statistics*, 2017, 99 (3), 435–448.
- Behrens, Kristian, Florian Mayneris, and Théophile Ndjanmou Biéda**, “Land for production: Evidence from Canadian manufacturing industries,” Technical Report, Université du Québec à Montréal 2020.
- Besag, Julian and James Newell**, “The detection of clusters in rare diseases,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1991, 154 (1), 143–155.
- Bleakley, Hoyt and Jeffrey Lin**, “Portage and path dependence,” *The Quarterly Journal of Economics*, 4 2012, 127 (2), 587–644.
- Bosker, Maarten, Eltjo Buringh, and Jan Luiten Van Zanden**, “From Baghdad to London: Unraveling urban development in Europe, the Middle East, and North Africa, 800–1800,” *Review of Economics and Statistics*, 2013, 95 (4), 1418–1437.
- Brooks, Leah and Byron Lutz**, “Vestiges of transit: Urban persistence at a microscale,” *The Review of Economics and Statistics*, 3 2019, 101 (3), 385–399.
- Bureau of Economic Analysis**, “Gross Domestic Product by industry data. Value added - 1947-2017: up to 71 Industries (Release 2018),” Technical Report 2018.
- Buringh, Eltjo and Centre for Global Economic History**, “Urbanisation Hub – The Clio-infra database on urban settlement sizes: 1500–2000,” Technical Report 2018.
- City Club of Chicago**, *The railway terminal problem of Chicago: A series of addresses before the City Club, June third to tenth*, City Club of Chicago, 1913.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon**, “The costs of agglomeration: House and land prices in French cities,” *The Review of Economic Studies*, 2018, 86 (4), 1556–1589.
- , —, —, **Diego Puga, and Sébastien Roux**, “The productivity advantages of large cities: Distinguishing agglomeration from firm selection,” *Econometrica*, 2012, 80 (6), 2543–2594.
- Daniels, P.W and A.M. Warnes**, *Movement in cities: Spatial perspectives on urban transport and travel*, 2nd ed., Routledge, 2007.
- Davis, Donald R and David E Weinstein**, “Bones, bombs, and break points: The geography of economic activity,” *American Economic Review*, 2002, 92 (5), 1269–1289.
- de Bellefon, Marie-Pierre, Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin**, “Delineating urban areas using building density,” *Journal of Urban Economics*, 2019, forthcoming.
- Dobbs, Richard, Sven Smit, Jaana Remes, James Manyika, Charles Roxburgh, and Alejandra Restrepo**, *Urban world: Mapping the economic power of cities*, The McKinsey Global Institute, 2011.
- Donaldson, Dave and Adam Storeygard**, “The view from above: Applications of satellite data in economics,” *Journal of Economic Perspectives*, 2016, 30 (4), 171–98.
- and **Richard Hornbeck**, “Railroads and American economic growth: A “market access” approach,” *The Quarterly Journal of Economics*, 2016, 131 (2), 799–858.

- Duranton, Gilles and Diego Puga**, "Urban land use," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 467–560.
- **and Matthew A. Turner**, "Urban growth and transportation," *The Review of Economic Studies*, 2012, 79 (4), 1407–1440.
- , **Peter M. Morrow, and Matthew A. Turner**, "Roads and trade: Evidence from the US," *Review of Economic Studies*, 2014, 81 (2), 681–724.
- Eckert, Fabian, Sharat Ganapati, and Conor Walsh**, "Skilled scalable services: The New Urban Bias in Recent Economic Growth," *University of California, San Diego. Mimeo*, 2020.
- Eeckhout, Jan and Philipp Kircher**, "Assortative matching with large firms," *Econometrica*, 2018, 86 (1), 85–132.
- Ellison, Glenn and Edward L. Glaeser**, "Geographic concentration in U.S. manufacturing industries: A dartboard approach," *Journal of Political Economy*, 1997, 105 (5), 889–927.
- Fujita, Masahisa**, "A monopolistic competition model of spatial agglomeration: Differentiated product approach," *Regional science and urban economics*, 1988, 18 (1), 87–124.
- **and Hideaki Ogawa**, "Multiple equilibria and structural transition of non-monocentric urban configurations," *Regional Science and Urban Economics*, 1982, 12 (2), 161–196.
- , **Paul Krugman, and Anthony Venables**, *The spatial economy: Cities, regions, and international trade*, The MIT Press, 1999.
- Gabaix, Xavier**, "The granular origins of aggregate fluctuations," *Econometrica*, 2011, 79 (3), 733–772.
- Galor, Oded and Ömer Özak**, "The agricultural origins of time preference," *American Economic Review*, 2016, 106 (10), 3064–3103.
- Gaubert, Cecile**, "Firm sorting and agglomeration," *American Economic Review*, 2018, 108 (11), 3117–53.
- Glaeser, Edward L.**, *Triumph of the city: How urban spaces make us human*, Macmillan, 2011.
- Godfrey, Brian J.**, "Urban development and redevelopment in San Francisco," *Geographical Review*, 1997, 87 (3), 309–333.
- Gonzalez-Navarro, Marco and Matthew A. Turner**, "Subways and urban growth: Evidence from earth," *Journal of Urban Economics*, 11 2018, 108, 85–106.
- Gyourko, Joseph and Raven Molloy**, "Regulation and housing supply," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 1289–1337.
- Harari, Mariaflavia**, "Cities in bad shape: Urban geometry in India," *American Economic Review*, 2020, 110 (8), 2377–2421.
- Haughwout, Andrew F.**, "Evidence from real estate markets of the long-term impact of 9/11 on the New York City economy," in Howard Chernick, ed., *Resilient city: The economic impact of 9/11*, Russel Sage Foundation, 2005, chapter 4, pp. 97–121.

- Heblich, Stephan, Stephen J. Redding, and Daniel M. Sturm**, "The making of the modern metropolis: Evidence from London," *The Quarterly Journal of Economics*, 2020, 135 (4), 2059–2133.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil**, "Measuring economic growth from outer space," *American Economic Review*, 2012, 102 (2), 994–1028.
- **and Arindam Mitra**, "The new urban landscape: Developers and edge cities," *Regional Science and Urban Economics*, 1996, 26 (6), 613–643.
- , **Tim Squires, Adam Storeygard, and David N. Weil**, "The global distribution of economic activity: Nature, history, and the role of trade," *The Quarterly Journal of Economics*, 2017, 133 (1), 357–406.
- Hornbeck, Richard and Daniel Keniston**, "Creative destruction: Barriers to urban growth and the Great Boston Fire of 1872," *American Economic Review*, 2017, 107 (6), 1365–1398.
- Krugman, Paul**, "Increasing returns and economic geography," *Journal of Political Economy*, 1991, 99 (3), 483–499.
- Lang, Robert E.**, *Edgeless cities: Exploring the elusive metropolis*, Brookings Institution Press, 2003.
- Lucas, Robert E. Jr. and Esteban Rossi-Hansberg**, "On the internal structure of cities," *Econometrica*, 2002, 70 (4), 1445–1476.
- Lui, Crocker H., Stuart S. Rosenthal, and William C. Strange**, "The vertical city: Rent gradients, spatial structure, and agglomeration economies," *Journal of Urban Economics*, 2018, 106, 101–122.
- Marshall, Alfred**, *Principles of economics*, London: Macmillan, 1890.
- McMillen, Daniel P.**, "Nonparametric employment subcenter identification," *Journal of Urban Economics*, 2001, 50 (3), 448–473.
- Michaels, Guy and Ferdinand Rauch**, "Resetting the urban network: 117–2012," *Economic Journal*, 2018, 128 (608), 378–412.
- , — , **and Stephen J. Redding**, "Task specialization in U.S. cities from 1880 to 2000," *Journal of the European Economic Association*, 2018, 17 (3), 754–798.
- Moyer, J. Alan**, "Urban growth and the development of the telephone: Some relationships at the turn of the century," in Ithiel de Sola Pool, ed., *The social impact of the telephone*, Cambridge, Massachusetts: MIT Press, 1977, pp. 318–348.
- Muller, Peter O.**, "Transportation and urban form: Stages in the spatial evolution of the American metropolis," in Genevieve Giuliano and Susan Hanson, eds., *The geography of urban transportation*, New York: Guilford Press, 2017, chapter 3, pp. 59–85.
- Nunn, Nathan and Nancy Qian**, "The potato's contribution to population and urbanization: Evidence from a historical experiment," *The Quarterly Journal of Economics*, 06 2011, 126 (2), 593–650.
- Philippon, Thomas**, "Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation," *The American Economic Review*, 2015, 105 (4), 1408–1438.
- PricewaterhouseCoopers UK**, "UK Economic Outlook (November 2009) – Which are the largest city economies in the world and how might this change by 2025?," 2009.

- Reba, Meredith, Femke Reitsma, and Karen C. Seto**, "Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000," *Scientific data*, 2016, 3, 160034.
- Redding, Stephen J. and Esteban Rossi-Hansberg**, "Quantitative spatial economics," *Annual Review of Economics*, 2017, 9 (1), 21–58.
- **and Matthew A. Turner**, "Transportation costs and the spatial organization of economic activity," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 1339 – 1398.
- , **Daniel M. Sturm, and Nikolaus Wolf**, "History and industry location: Evidence from German airports," *The Review of Economics and Statistics*, 8 2010, 93 (3), 814–831.
- Ripley, Brian D**, *Spatial statistics*, Vol. 575, John Wiley & Sons, 2005.
- Roseman, C.C., R. Wallach, and D. Taube**, *The Historic Core of Los Angeles Images of America*, Arcadia, 2004.
- Rosen, Sherwin**, "Authority, control, and the distribution of earnings," *Bell Journal of Economics*, 1982, 13 (2), 311–323.
- Rozenfeld, Hernán D., Diego Rybski, Xavier Gabaix, and Hernán A. Makse**, "The area and population of cities: New insights from a different perspective on cities," *American Economic Review*, August 2011, 101 (5), 2205–25.
- Santamaria, Marta**, "The Gains from Reshaping Infrastructure: Evidence from the division of Germany," *University of Warwick. Mimeo*, 2020.
- Sassen, Saskia**, *The global city*, Princeton: Princeton University Press, 2002.
- Sequeira, Sandra, Nathan Nunn, and Nancy Qian**, "Immigrants and the making of America," *The Review of Economic Studies*, 3 2019, 87 (1), 382–419.
- Siodla, James**, "Clean slate: Land-use changes in San Francisco after the 1906 disaster," *Explorations in Economic History*, 2017, 65 (April), 1–16.
- Taylor, Peter J**, "Leading world cities: empirical evaluations of urban nodes in multiple networks," *Urban studies*, 2005, 42 (9), 1593–1608.
- Trujillo, Jesus Leal and Joseph Parilla**, *Redefining global cities*, The Brookings Institution, 2016.

A Appendix material

A.1 Importance of prime services in the US

Technical definition and national importance of prime services. As a first approximation, one can define *prime services* in a national accounting sense as the sum of the ‘Finance, insurance, real estate, rental, and leasing’ (FIRE; NAICS 52-53) and ‘Professional and business services’ (PROF; 54-56) sectors in the national accounts.

Table A 1: Prime Services in National Accounts (2016)

Sub-sector	Sub-sub-sector	% of GDP	NAICS	BEA
<i>FIRE: Finance, insurance, real estate, rental, and leasing</i>				
Finance and insurance			52	52
	Federal Reserve banks, credit intermediation, and related activities	2.87	521, 5221, 5222-3	521CI
	Securities, commodity contracts, and investments	1.26	5231-2, 5239	523
	Insurance carriers and related activities	3.15	5241, 5242	524
	Funds, trusts, and other financial vehicles	0.26	525	525
Real estate and rental and leasing				531
	Real estate - Housing	9.90	531	5310HS
	Real estate - Other real estate	2.31	531	531ORE
	Real estate - Rental and leasing services and lessors of intangible assets	1.09	532	532RL
<i>PROF: Professional and business services</i>				
Legal services		1.32	5411	5411
Computer systems design and related services		1.50	541511, 541512, 541513, 541519	5415
Miscellaneous professional, scientific, and technical services		4.30	5412, 5413, 5414, 54161, 54162, 54169, 5417, 5418, 54191, 54193, 54199 54192 54194	5412OP
Management of companies and enterprises		1.92	55	55
Administrative and waste management services		3.05	56	561
Σ		28.80		

Notes: Data are from the Sectoral National Accounts of the United States (Bureau of Economic Analysis, 2018). Rows with sectors including prime services are colored gray. The NAICS codes refer to the 2007 classification.

However, these broad categories also contain other services, for example leasing services and facility management, which are arguably not prime services in our definition. We thus further disaggregate the FIRE and PROF sectors (Table A 1). The exclusion of “Administrative and waste management services” and “Real estate - Rental and leasing services and lessors of intangible assets” reduces the total share of prime services from about 33% to about 29%. Furthermore, the disaggregation highlights the importance of real estate among

prime services (around 12.2% of total value added). Yet, even if we exclude the real estate sector, the share of prime services in total value added today remains significantly larger than that of manufacturing. Likewise, the pattern of its evolution since the 1940s remains virtually unchanged (corresponding graph available upon request).

Geographic concentration of prime services in MSAs. The Bureau of Economic Analysis (BEA) provides industry-level value added GDP estimates for the 382 Metropolitan Statistical Areas (MSAs) of the United States. Table A 2 shows that in terms of their contribution to national GDP, prime services are over-represented in metropolitan areas relative to non-metropolitan areas.

Table A 2: Prime Services in National Accounts in 2015

Prime service	MSAs		Non-MSAs		Total M\$
	M\$	Share	M\$	Share	
Finance and insurance	1,237,398	95.69%	55,702	4.31%	1,293,100
Real estate	2,054,375	95.21%	103,425	4.79%	2,157,800
Professional, scientific, and technical services	1,252,097	96.85%	40,703	3.15%	1,292,800
Management of companies and enterprises	345,625	97.03%	10,575	2.97%	356,200
All prime services	4,889,495	95.87%	210,405	4.13%	5,099,900
In comparison:					
Manufacturing	85.65%		14.35%		
National GDP	90.26%		9.74%		

Notes: Own calculation based MSA GDP release ‘Gross Domestic Product by Metropolitan Area, Advance 2016, and Revised 2001-2015’ (September 20, 2017 release: [download here](#)). These data are fully consistent with the national accounts, which allows us to decompose the value-added GDP and its components into an MSA and a non-MSA part. The corresponding National GDP data release is ‘Gross Domestic Product by Industry and Input-Output Statistics’ (November 3, 2016 release: [download here](#)).

To estimate the importance of the urban bias between metro areas of different sizes (Figure 1), we run the following regressions:

$$\frac{VA_{sm}}{VA_m} = c_s + a_s \ln P_m + e_{sm},$$

where VA is value added, $s \in \text{manufacturing, prime services, all other sectors}$ indexes sectors, and m indexes metro areas. P_m is the 2010 metro population, a_s is the semi-elasticity of interest, and e_{sm} is an error term. We estimate this model at the sector- s level.³⁰ Figure 1 shows the results for prime services.

³⁰For some metropolitan areas, fine-grained industry-specific data are not released because of privacy concerns. It is thus necessary to impute the shares of certain industries using higher-level aggregates. We discuss this procedure, reference the relevant population sources, and show the robustness of our results towards excluding the imputed data in our Global Cities dataset appendix (Ahlfeldt et al., 2020b). As expected, dropping the imputed data increases the elasticity of interest as places with a low prime service density are more prone to not having data released.

A.2 Estimating employment weights for prime points

This section describes how we assign employment weights to prime points (PPs), and how we overidentify the employment-weighted prime points. In a nutshell, we assign employment weights to distinct types of prime points using establishment-level datasets which we collect for six cities (Boston, Montreal, New York, Philadelphia, Toronto, and Vancouver). We then overidentify the established relationship using data on Starbucks, prime office, and co-working spaces at the establishment level as well as zip-code CBP data for 39 US cities from our dataset that do not serve to compute those weights.³¹

A.2.1 Estimating weights

To take full advantage of the spatial granularity of our datasets—all observations are geocoded by latitude and longitude coordinates—we employ a point-pattern-based empirical approach. For each city, we generate a set of $l \in L$ locations that consist of disks with a radius of 750 meters (about a square mile) drawn around random points sampled from within the cities. We discard locations sampled in undevelopable parts of the city (e.g., water areas, steep slope) and adjust our disks for the share of undevelopable land in them. Hence, the predictive power of prime points is not driven by some areas being undevelopable. For each location l drawn in city c , we compute its aggregate prime service employment $\mathcal{E}_{l,c}$ from our establishment-level data, as well as a separate counts $\mathcal{P}_{l,c}^T$ for each type of prime point T (e.g., accounting firms, law firm) in the disk. As discussed in Section 2.2, we create separate sub-categories for establishments of globally operating companies (e.g., PWC, Deloitte). To establish an empirical link between $\mathcal{E}_{l,c}$ and $\mathcal{P}_{l,c}^T$, we estimate the following empirical specification in city-specific regressions:

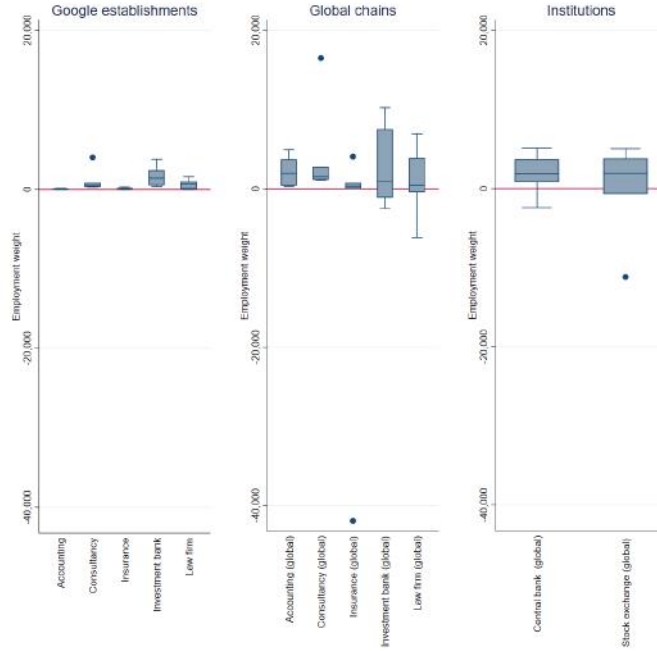
$$\mathcal{E}_{l,c} = a + \mathcal{P}_{c,l}^T b_c^T + \varepsilon_{l,c}^T \quad (15)$$

where b_c^T are city-establishment-type-specific employment weights and $\varepsilon_{l,c}^T$ is a residual term. Figure A1 provides an accessible presentation of the results: it shows box plots of the distribution of employment weights across cities by establishment type. The vast majority of employment weights is positive.³² As expected, establishments by global leader companies generally receive much larger weights than their non- ‘global leader’ counterparts. Considering that we employ simple and transparent multivariate regression models, the predictive power of our prime points is quite impressive. Generally, we explain more than 90% of the variation. Philadelphia is the exception, but we still explain close to 80% of the variation.

³¹We document the sources of the underlying data (estimation: NETS data, Scott’s data, scraped prime points; validation: point data on SNL, Starbucks, co-working spaces, and ZIP/ZCTA-level data from County Business Patterns) in the accompanying Global Cities dataset appendix (Ahlfeldt et al., 2020b).

³²There are a small number of significant negative coefficients, which may be due to substantial collinearity between the variables.

Figure A1: Employment weights by city and establishment type



Notes: Box plot illustrates the distribution of employment weights across cities by establishment type. Employment weights are from city-specific regressions of observed prime services employment at the establishment-level (NETS or Scott’s data) on prime points (queries from the Google Places API and scraped from company websites).

Table A 3: Median employment weights for EWPPs

Establishment type	Weight	Establishment type	Weight
Global chain: Accounting	1898	Google places: Accounting	68
Global chain: Consultancy	1507	Google places: Consultancy	450
Global chain: Insurance	314	Google places: Insurance	98
Global chain: Investment bank	918	Google places: Investment bank	1406
Global chain: Law firm	409	Google places: Law firm	676
Stock exchanges	1879		
Central banks (incl. branches)	1846		

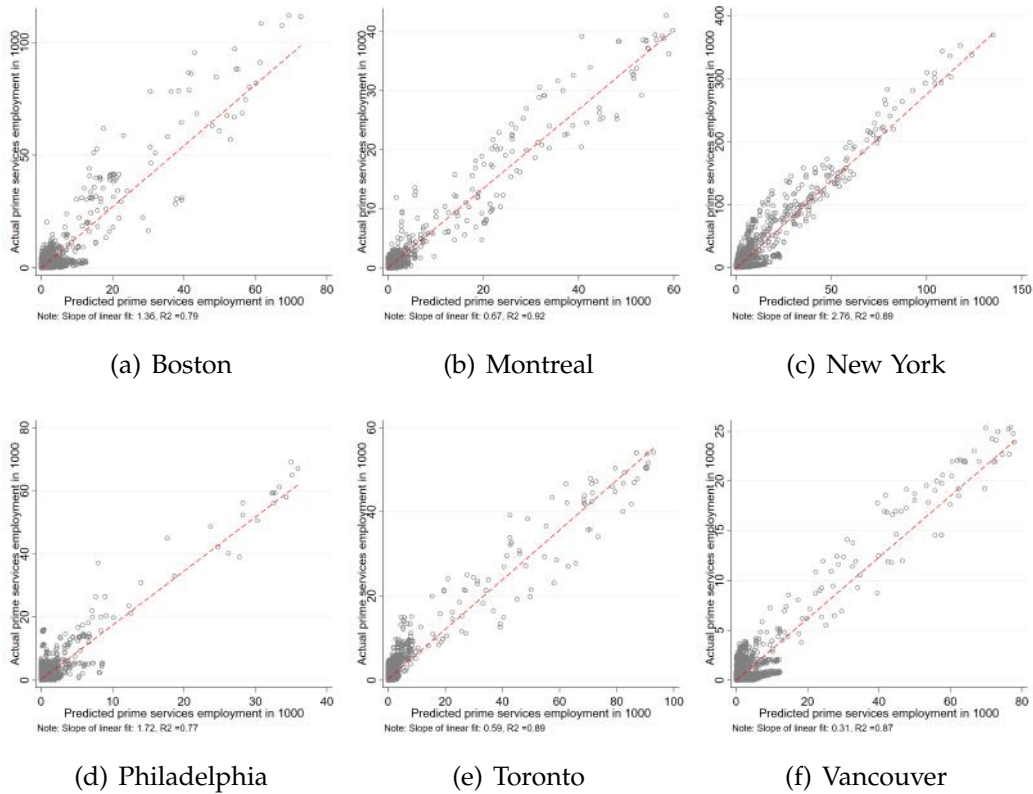
Notes: Employment weights from city-specific regressions of PS employment on PP counts (within random 750-m-radius disks). The values reported in the table are the median values in the distributions of estimated employment weights across cities by establishment type. Values are rounded to the nearest integer so that predicted employment figures are integers, too.

Since we wish to use the estimated employment weights \hat{b}^T for extrapolation to other cities, we need to abstract from city-specific heterogeneity and settle on PP-type-specific weights. We opt for the median of the distribution across cities as the obvious choice to avoid that the PP-type-specific weights are driven by outliers. Table A 3 shows the median weights that we obtain, which are all positive.

It is straightforward to use \hat{b}^T to generate a measure of predicted prime service employment $\hat{\mathcal{E}}_{l,c} = \mathcal{P}_{c,l}^T \hat{b}^T$. Since our PPs consist of establishments that employ prime service workers, the associated employment weights have a very intuitive interpretation: they provide a type-specific proxy for unobserved employment at the PPs. Figure A2 shows scatter-plots of

the correlation between $\widehat{\mathcal{E}}_{l,c}$ and $\mathcal{E}_{l,c}$ by city. Although we use the median weights \widehat{b}^T which do not vary by city in the construction of $\widehat{\mathcal{E}}_{l,c}$, the correlations are still strong. The R^2 from simple bivariate linear regressions ranges from .77 to .92 for all cities. Given our focus on the densest clusters of prime services, it is particularly reassuring that we reliably predict where prime services density is highest.

Figure A2: Prime services vs. prime point densities



Notes: Figure compares the observed prime service employment to the predicted one at the level of 30,000 750-radius-disks drawn around 30,000 random points.

A.2.2 Overidentification

Given their crucial role in our analysis, we now provide a battery of overidentification tests for our EWPP measures. To do so, we rely on *prime places* that are closely linked to prime service production. SNL-S&P investments typically comprise grade-A office buildings, which are the natural places of prime service production. Starbucks franchises are arguably the most universal workplace amenity and places where workers in knowledge-based tradable services can meet to interact. Co-working spaces are a mix of both. Hence, we expect our prime service employment measure to be a relevant predictor of the presence of SNL-S&P buildings, Starbucks franchises, and co-working spaces. The regressions in columns (1)–(4) of Table A 4 are out-of-sample overidentification tests since no information on the dependent

variables has been used in the construction of our predicted employment measure. Our predicted prime services employment measure is a strong predictor of the spatial distribution of prime places (SNL-S&P buildings, Starbucks franchises, and co-working spaces), thus suggesting external validity within the sampled cities.

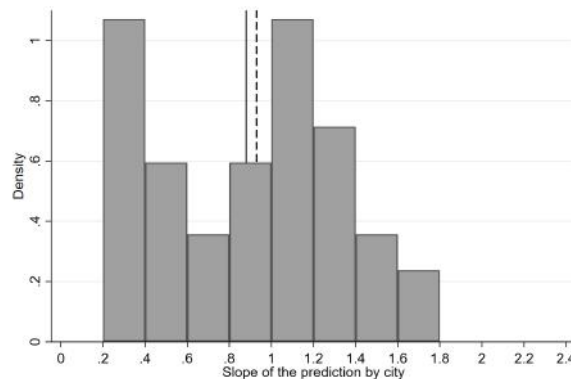
Table A 4: Overidentification of EWPPs

	(1)	(2)	(3)	(4)	(5)	(6)
	Prime places count	SNL-S&P count	Starbucks count	Co-working count	Actual PS emp. (1000)	Actual PS emp. (1000)
Predicted PS employment (1000)	0.928*** (0.016)	0.371*** (0.005)	0.351*** (0.014)	0.205*** (0.004)	1.007*** (0.165)	0.493*** (0.040)
Constant	0.091*** (0.008)	0.121*** (0.003)	0.010 (0.006)	-0.033*** (0.002)	0.067 (0.281)	0.997*** (0.073)
Spatial unit	750-m radius disks	750-m radius disks	750-m radius disks	750-m radius disks	Zip/ZCTA level	Zip/ZCTA level
Cities	NETS	NETS	NETS	NETS	US NETS	US NON-NETS
N	131346	131668	131668	131346	1145	4929
R ²	.687	.616	.418	.693	.411	.337

Notes: Unit of observation varies according to the indicated spatial unit. Prime places count is the sum of SNL-S&P, Starbucks, and co-working counts. Robust standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Next, we use our employment weights and PPs to predict prime service employment at the level of US zip code tabulation areas. At this level, we can merge our predicted prime service employment measure to prime service employment as recorded in the US County Business Patterns (CBP). In column (5), we regress CBP prime service employment on our predicted employment at the zip-code level for the three US cities (Boston, New York, and Philadelphia) for which we have used micro-geographic data in the construction of the weights. We explain 40% of the variation and estimate a positive slope coefficient that is highly statistically significant but insignificantly different from one.

Figure A3: Actual vs. predicted employment by non-targeted city



Notes: Histogram shows the distribution of slope parameters from city-specific regressions of actual (from CBP) to predicted (EWPPs) prime services employment at the level of US zip code tabulation areas.

In column (6), we push the idea of an overidentification test further. We now use the weights estimated for Boston, New York, and Philadelphia to predict prime service employment for the other 39 US cities in our global sample that do not overlap with these three

cities. While this is a particularly demanding overidentification test, we still explain more than one third of the variation within and between cities.³³ The slope coefficient is about 0.5, which implies that the establishments we use to predict prime service employment are, on average, about half as large in the US cities used for overidentification than in the three cities we use for estimating the weights. Figure A3 plots the histogram of the predicted slope coefficients by city for the 39 cities. The relatively small coefficient is driven by a relatively large number of cities where prime establishments tend to be small. The median and mean slope across US cities, at 0.9, is much closer to 1.

A.3 Validation of the clustering algorithm

Table A 5 shows how prime locations specialize in prime services, relative to their host cities. Zooming into the six US and Canadian cities, for which we observe establishment-level employment, we find a striking specialization of prime locations in prime services. The share of prime service employment within 750 meters of any prime location ranges from a low of 42% (Philadelphia) to a high of 68% (Toronto). The average prime service share, at 55.8%, exceeds the manufacturing share by a factor of 17. For global prime locations, this ratio increases to more than 22. Our prime locations are thus way more specialized in prime services than their host cities for which the average shares, at 26.5% (prime services) and 17.1% (manufacturing), are more similar (see Table A 5). As anticipated, this effect is even stronger for global prime locations. In sum, our (global) prime locations pick up substantial concentration in prime services as measured from detailed micro-geographic data.

Table A 5: Prime location specialization

	NYC	Boston	Philly	Toronto	Montreal	Vancouver	Mean
Within 50 km of	Empire State	Prudential	Liberty Bell	CN Tower	KPMG Tower	MNP Tower	Any CBD
Share MFG	7.48%	9.38%	8.42%	27.26%	27.72%	22.36%	17.10%
Share PS	28.48%	27.50%	26.08%	26.78%	24.26%	26.10%	26.53%
Metro PS/MFG ratio	3.81	2.93	3.10	0.98	0.88	1.17	2.15
Within 750m of global PL							
Share MFG	5.65%	1.54%	3.86%	2.37%	2.95%	1.86%	3.04%
Share PS	59.43%	60.73%	43.13%	68.81%	60.94%	50.47%	57.25%
Global PL PS/MFG ratio	10.52	39.44	11.17	29.03	20.66	27.13	22.15
Within 750m of PL							
Share MFG	5.16%	2.57%	3.56%	3.12%	2.95%	3.36%	3.45%
Share PS	58.99%	55.08%	42.21%	67.65%	60.94%	49.73%	55.77%
Any PL PS/MFG ratio	11.43	21.43	11.86	21.68	20.66	14.80	16.98

Notes: MFG = manufacturing; PS = prime services; PL = prime locations. Employment data for New York, Boston, and Philadelphia from the National Establishment Timeseries database (NETS, 2012). Employment data for Toronto, Montreal, and Vancouver from the Scott's National All database for the year 2013. For details, see the Global Cities dataset appendix (Ahlfeldt et al., 2020b). MFG is delimited by NAICS 31–33. See Appendix A.1 for the NAICS-codes of PS. All missing employment figures have been replaced with the median establishment employment (for either MFG, PS, or all establishments). For selection of prime locations, see Section 2.4.

³³Dropping the ZCTAs without prime service employment hardly changes the results. If we add city fixed effects to the regression in column (6) of Table A 4, the R^2 increases to close to 40%.

A.4 Measures of spatial concentration

Table A 6: Measures of spatial concentration

Measure	Shade of primeness	Formula	Notation & comments
Modern concentration measures			
Bilateral distances between...			
...EWWP (weighted)	General prime service sprawl	$G = \frac{1}{\sum_i \sum_{k \neq i} E_i E_k} \sum_i \sum_{j \neq i} d_{ij} E_i E_j$	d_{ij} : bilateral distance between two prime points i and j , $j \neq i$, and E is an employment weight.
...Prime Locations (unweighted)	Gravity of PLs among each other	$G = \frac{1}{\sum_i \sum_{k \neq i} E_i E_k} \sum_i \sum_{j \neq i} d_{ij} E_i E_j$	i and j are prime locations; the average is unweighted: $E_i = E_j = 1$
Cumulative densities	Overall compactness	$K_m(d) = \frac{1}{hN(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N f\left(\frac{d_{ij}-d}{h}\right)$	d_{ij} denotes the great circle distance between establishments i and j ; h : optimal bandwidth (set using Silverman's rule); f : Gaussian kernel.
Central point	Direction (N/E/S/W) of center relocation (visual)	Concentration relative to early center $\arg \min_i G_i = \frac{1}{\sum_{k \neq i} E_i E_k} \sum_{j \neq i} d_{ij} E_i E_j$	d_{ij} : bilateral distance between two prime points i and j , $j \neq i$, and E is an employment weight.
Distance density gradient	Monocentricity of historic core today	$\ln(n_{dm}) = \alpha_m + \beta_m \ln(D_{dm}) + \epsilon_{dm}$	n_{dm} : establishment density within distance rings d with radii 0-250, 250-750, 750-1250 etc. D_{dm} : distance of a distance bin. β_m : metro-specific distance gradient (measure of concentration).
Distance from historic centers for...			
...EWWP	Gravity of old center for establishments	$\overline{d}_m = \frac{1}{N_m} \sum_i d_{im}$	m : city hall 1900; i : EWWP location
...all PLs	Attraction of old center for new centers	$\overline{d}_m = \frac{1}{N_m} \sum_i d_{im}$	m : city hall 1900; i : prime locations
...global PLs	Persistence of most important centers	$\overline{d}_m = \frac{1}{N_m} \sum_i d_{im}$	m : city hall 1900; i : global prime locations

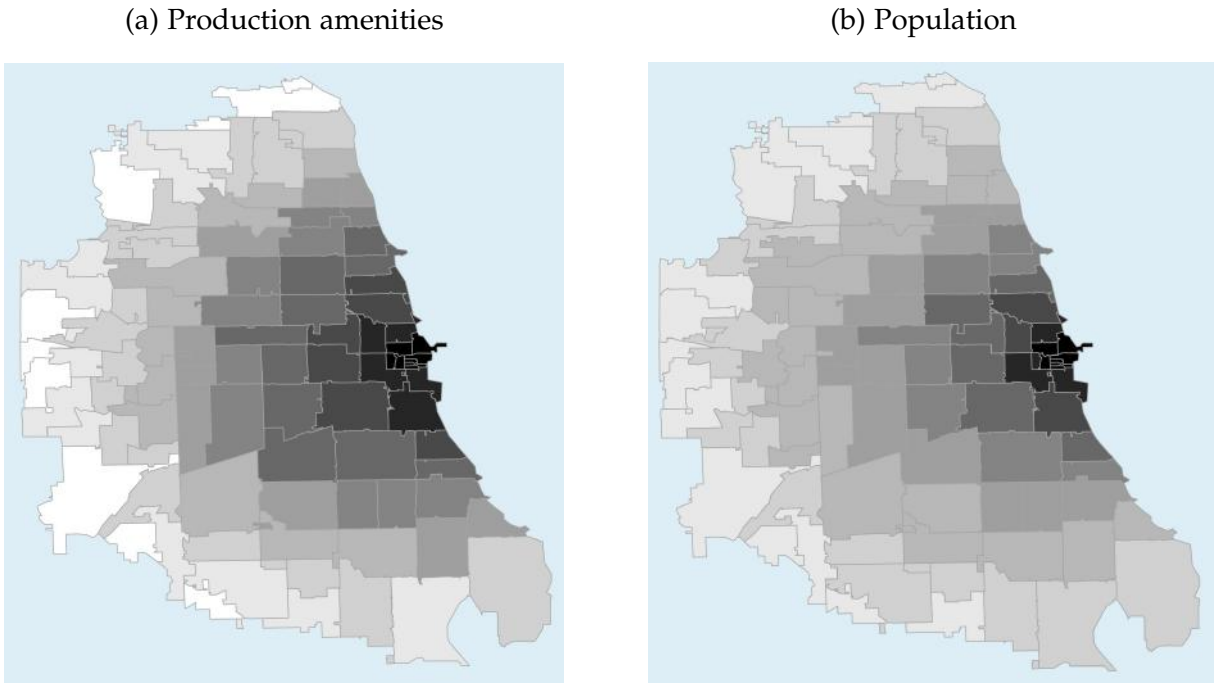
A.5 ABM parametrization and model output

In this appendix, we provide additional information on the parametrization of the model and additional output to illustrate a typical model run.

A.5.1 Geography

We simulate our model using ZCTAs and a simplified version of the city of Chicago. We restrict ourselves to a radius of 15 miles around the CBD (as defined on the City of Chicago data portal), which is similar to the spatial scale we use for our empirical analysis (panel (b) of Figure 8 shows the extent of our empirical grid).

Figure A4: Distribution of production amenities and population



Notes: Exogenous productivity gradient and sample initial allocation of population in the ABM.

ZCTA 60601 is the historic center that contains the foundation place (FP) from our urban biographies data set. We set productivity $B_{FP} = 1.3$ and generate a gradient of decreasing productivity from that center: $B_\ell = e^{-0.01 \times d_{\ell,FP}} \times B_{FP}$. Hence, the historic foundation place has a productivity advantage over places that are farther away. Note, however, that the exogenous productivity differences are small, contrary to quantitative spatial models where unobserved productivity differences across space are usually huge. We further allocate population and the initial distribution of firms using the following distance-based shares:

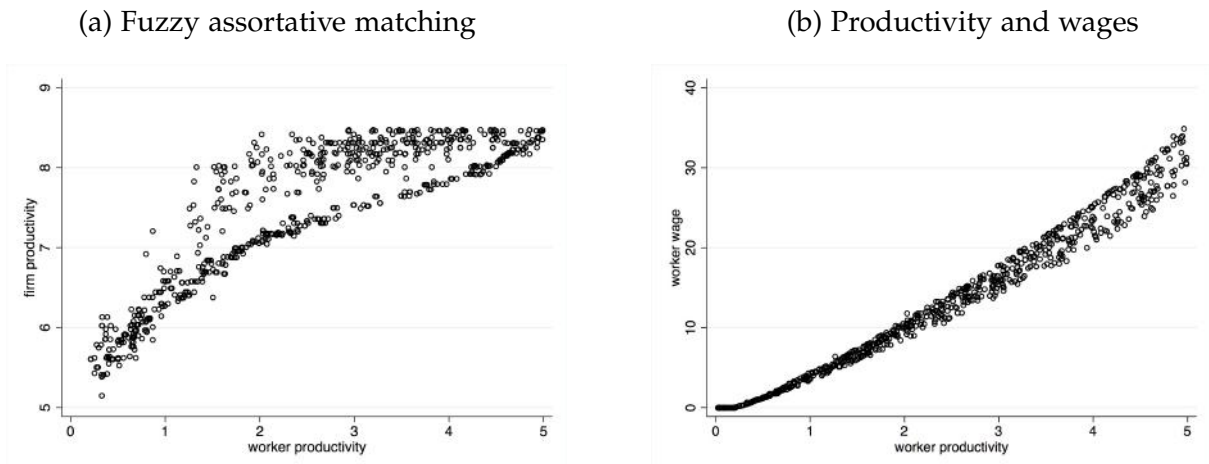
$$\text{pop}_\ell = \frac{e^{-0.1 \times d_{\ell,FP}}}{\sum_\zeta e^{-0.1 \times d_{\zeta,FP}}} \times \text{total pop} \quad \text{and} \quad \text{firms}_\ell = \frac{e^{-0.4 \times d_{\ell,FP}}}{\sum_\zeta e^{-0.4 \times d_{\zeta,FP}}} \times \text{total firms}. \quad (16)$$

Since agents are indivisible in the ABM, we round the results to the nearest integer. We set the number of firms to be $1/3$ the number of the total population, again rounded to the nearest integer.³⁴ Panel (a) of Figure A4 shows the spatial structure of our production amenities B_ℓ , while panel (b) displays the allocation of workers. This initial allocation is fairly monocentric.

A.5.2 Heterogeneous agents

Workers and firms are heterogeneous. They each get a productivity draw θ_w (workers) or θ_φ (firms) from a uniform distribution. Firms draw from $\theta_\varphi \in [5, 8.5]$, whereas workers draw from $\theta_w \in [0, 5]$. Firms' bargaining power is set to 0.33, i.e., one-third of revenue goes to profits and two-thirds of revenue goes to wages. The span-of-control parameter of each firm is a uniform random draw between 0 and 4; the SOC elasticity is set to 1.4 for all firms. Each firm requires 2.1 units of office space per worker.³⁵ We impose no correlation between θ_φ and \bar{R}_φ , i.e., firms differ along two independent dimensions in their productivity.

Figure A5: Fuzzy assortative matching, productivity, and wages



Notes: Sample run of the ABM. The left panel depicts the relationship between firm and worker productivity and shows 'fuzzy' assortative matching. The right panel depicts the relationship between worker productivity and wages.

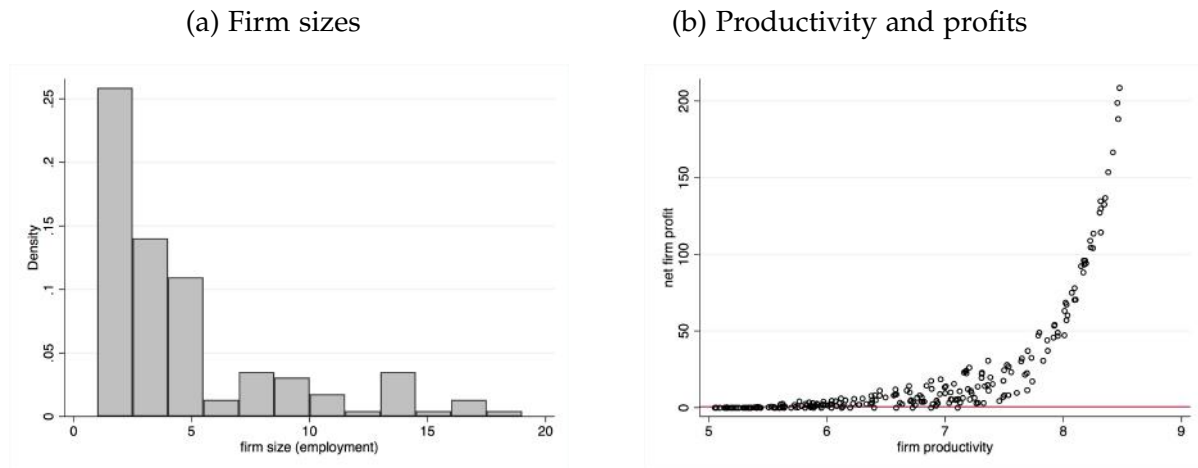
We sort workers and firms by productivity and run an initial assignment where the most productive workers choose their matches first. The initial assignment has 186 active firms and 671 assigned workers. Hence, 25 workers remain unemployed (get zero wage) and 43 potential firms (seeds) cannot attract workers (end up being inactive). Obviously, by the

³⁴The total population is set at total pop = 700 in our illustration. Given rounding, this implies an initial allocation of 696 workers and 229 (potential) firms. In the Monte Carlo simulations, we randomly draw a population size between 500 and 1000 from a uniform distribution.

³⁵The rationale underlying these parameter choices is that we want sufficient variation in firm sizes without having only a very small number of giant firms.

matching process it is the low-productivity workers who remain unemployed (average productivity of 0.101, compared to 2.536 for the employed workers); and the low-productivity firms that do not operate (average productivity of 5.576, compared to 6.963 for the operating firms). More productive firms attract (on average) more productive workers and are larger. However, since there are two dimensions of heterogeneity (productivity and the SOC rank), the relationship is fuzzy (see panel (a) of Figure A5). Assortative matching leads (on average) to a convex relationship between productivity and wages (see panel (b) of Figure A5).

Figure A6: Firm size and profit distributions



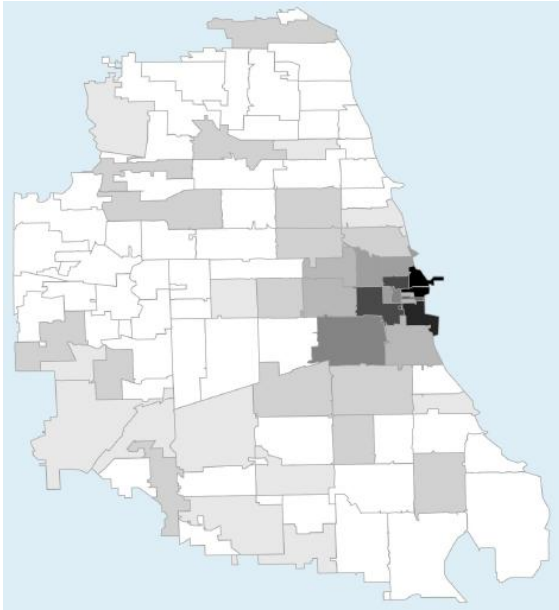
Notes: Sample run of the ABM. The left panel provides a histogram of firm sizes (employment), whereas the right panel displays profit by firm productivity. Inactive firms make zero profit.

Turning to the firm-size distribution, it is depicted by panel (a) of Figure A6. The matching process and our parameterization yield a distribution of firm sizes with many small firms and a few large players, a realistic feature of observed firm-size distributions that are close to Pareto. A simple regression of log rank on log size yields a slope of -1.16 (standard error 0.035 and adjusted R^2 of 0.856) in the initial, and -1.04 (standard error .035 and adjusted R^2 of 0.853) in the final equilibrium. The skewness in firm sizes maps into skewed profits (see panel (b) of Figure A6). The aggregate value of output produced in the example is 15,162, of which 5,003 (one third) go to firms' profits; and 10,158 (two thirds) are paid to workers. Workers' income net of commuting costs is 9323, i.e., about 85% of their gross income. Thus, aggregate commuting costs are about 15% of gross income, a reasonable number.³⁶

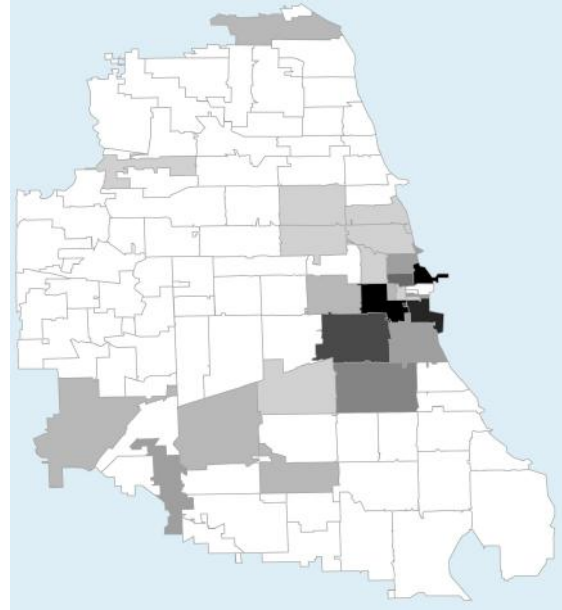
³⁶According to BLS data for 2019, US consumers spent on average \$63,036 ('Average annual expenditures') of which \$10,742 was dedicated to commuting and other trips ('Transportation') (see <https://www.bls.gov/news.release/cesan.nr0.htm> for the data). This represents about 17% of income. The 2017 figure stands at 15.9%.

Figure A7: Initial and final distributions of employment

(a) Initial employment



(b) Final employment



Notes: Initial and final equilibrium distribution of employment. Darker colors mean more employment, white means zero employment.

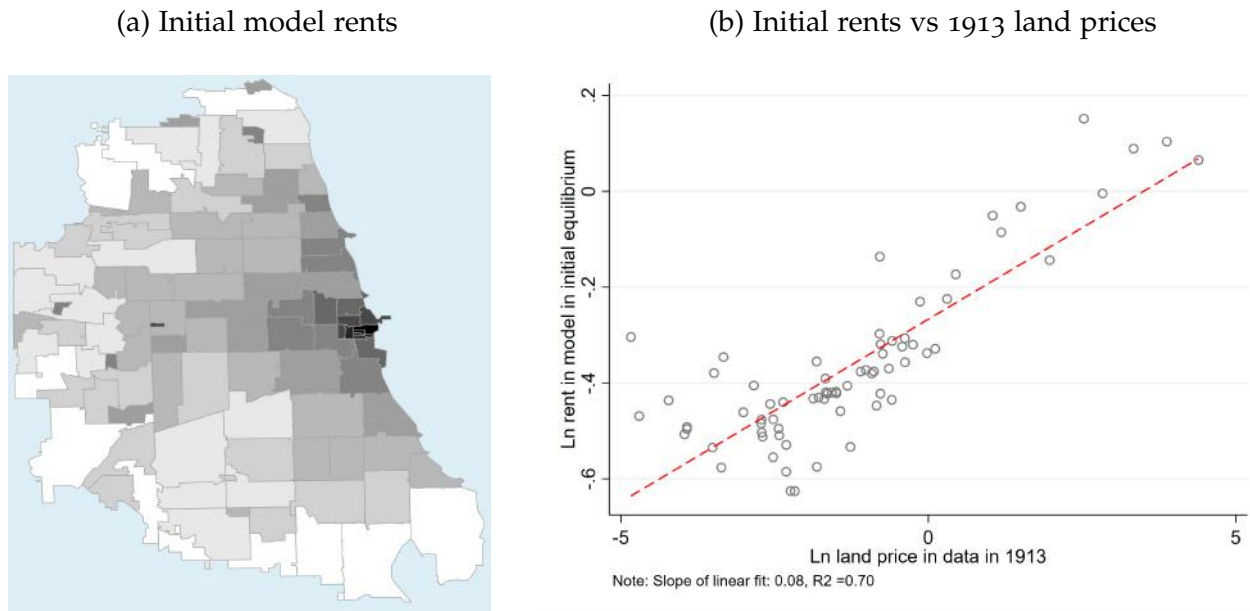
A.5.3 Equilibrium

An equilibrium is such that: (i) workers optimally choose firms; (ii) firms optimally choose locations;³⁷ and (iii) rents clear the real estate market in all locations. The initial equilibrium distribution of employment is shown in panel (a) of Figure A7. It is fairly monocentric and concentrated in the historic center of the city (where the productivity advantage is the largest). The initial equilibrium has one prime location that contains 48.8% of the total employment in our model city. Observe also that more than half of the locations have zero equilibrium employment. Panel (a) of Figure A8 depicts the corresponding spatial structure of land rents in the initial equilibrium. As shown in panel (b) of Figure A8, this allocation mimics well the qualitative structure of the 1913 Chicago land price gradient.

We allow for a transport network to be potentially developed after the initial equilibrium is established. We assume a sunk setup cost of 10,000; a per-station cost of 200; and a per kilometer cost of 50. A uniform tax of 15% is collected on wages and profits in the initial equilibrium to build stations and links while budget is available. More details on the algorithm for network construction—and on a number of choices we made—is given in a technical appendix (Ahlfeldt et al., 2020a). In this example, a network with 20 stations

³⁷For numerical reasons we impose small moving costs for firms (equal to 0.5 in this case) to avoid limit cycles where firms try to arbitrage away tiny profit differences (recall that everything is indivisible in this model). This does not drive our results and introduces some additional randomness into the initial spatial allocation.

Figure A8: Initial rents vs historic land prices in 1913



Notes: Model land rents (real estate prices) in the initial equilibrium and historic land values in Chicago, 1913 (data from [Ahlfeldt and McMillen, 2018](#)).

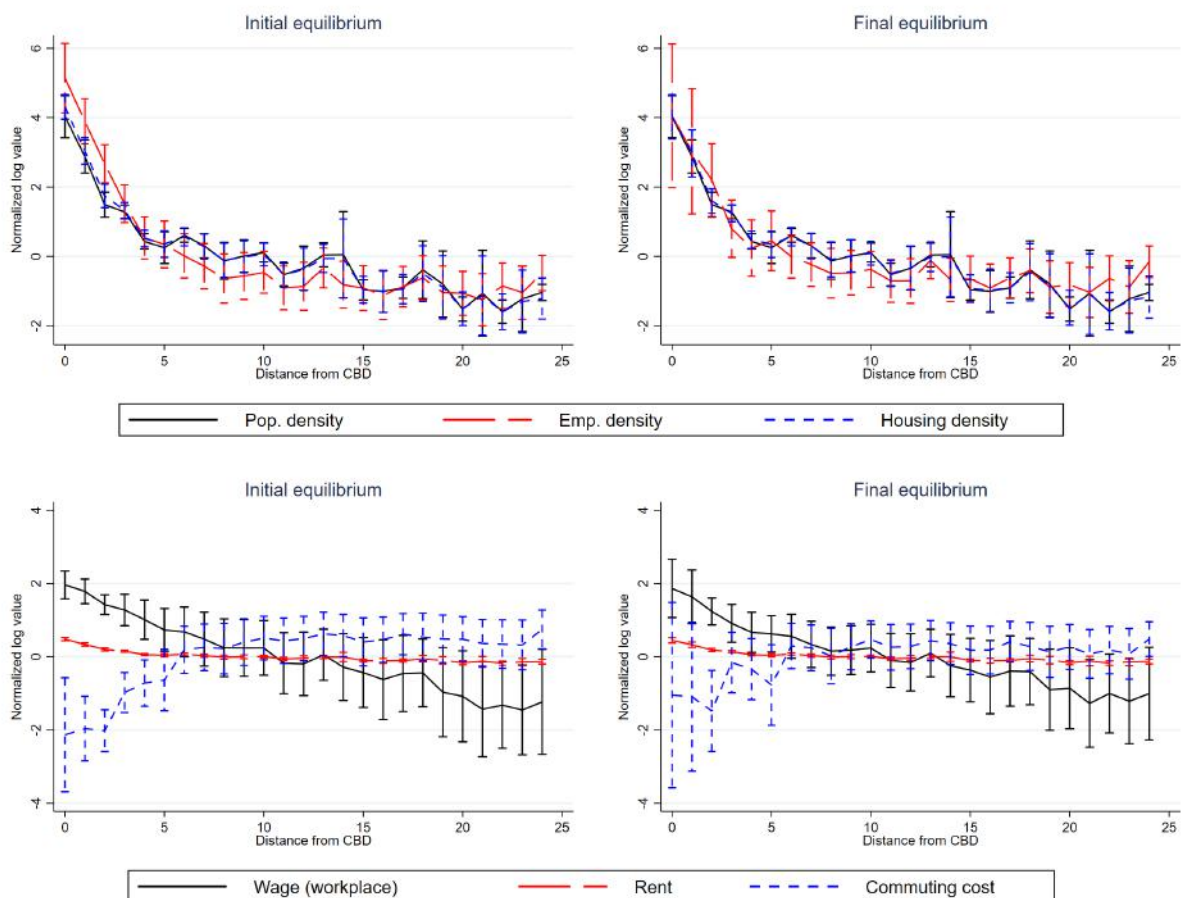
and total length of 31.34 kilometers is developed. Panel (b) of Figure 8 depicts the network, with the prime location in the initial equilibrium in red. Since employment is highly concentrated in the city, the commuting flow matrix is dominated by zeros. In our initial equilibrium, 96.6% of the 10,404 ($= 102 \times 102$) bilateral commuting flows are zero. Thus, 5.9% of destinations account for 70.3% of the commutes. This provides a strong rationale for the development of a hub-and-spokes type of network centered on the CBD, thus providing another endogenous locational advantage to the CBD.³⁸ We finally also estimate a standard commuting gravity regression in the initial and the final equilibrium that controls for arbitrary origin and destination effects in each simulation run. Commuting declines at a rate of about 5% per km, which assuming an average speed of 30km/h, closely matches the decay observed in real-world commuting data ([Ahlfeldt et al., 2015](#)).

Turning to the final equilibrium, we move to a second-nature world and allow for agglomeration economies and shocks. We assume $\epsilon = 0.04$ for employment in the own ZCTA, and $\epsilon_n = 0.01$ for employment in neighboring ZCTAs (where neighbors are ZCTAs with centroids up to 2 kilometers away). Our values are in line with recent estimates of agglomeration effects from detailed micro-data (see [Combes and Gobillon 2015](#) for a survey). In our illustrative example, the initial prime location gets hit by a shock that forces a subset of firms—accounting for 52.74% of the PL employment—to move. This leads to a new

³⁸We assume that the cost of moving on the network is $\zeta = 0.3$ times the cost of moving off the network. In other words, moving 1km on the network is equivalent to moving 0.3km off the network. Of course, one has to access the network first, which may be impractical and thus make the network a less preferred choice.

short-run equilibrium with two prime locations totaling 173 and 289 employees, respectively. Once the shock dissipates and all firms re-optimize, the larger PL absorbs the smaller one and grows to 538 employees (79.94% of employment). Observe that geographic concentration is higher in the final equilibrium. This is normal because we move from a first-nature world without externalities to a second-nature world with externalities. Yet, our simulations and regression analysis reveal that cities which experience larger shocks and/or do not develop a transportation network (or develop a smaller network) tend to be less spatially concentrated in the final equilibrium than other cities.

Figure A9: Various distance gradients in the ABM

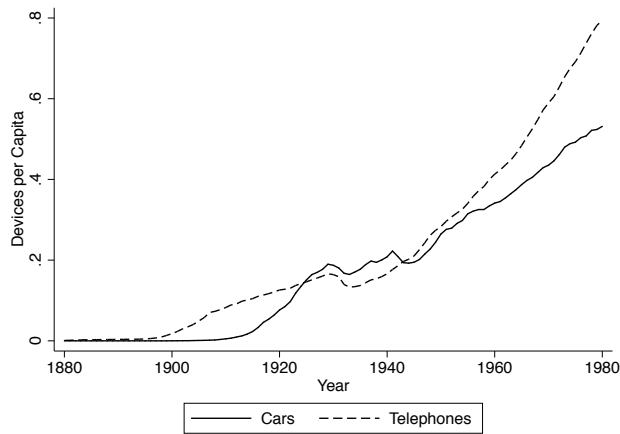


Notes: CBD is the centroid of the block that is the prime location in the majority of runs. Lines connect averages across blocks within 1-mile distance bins with error bars giving +/- one standard deviation.

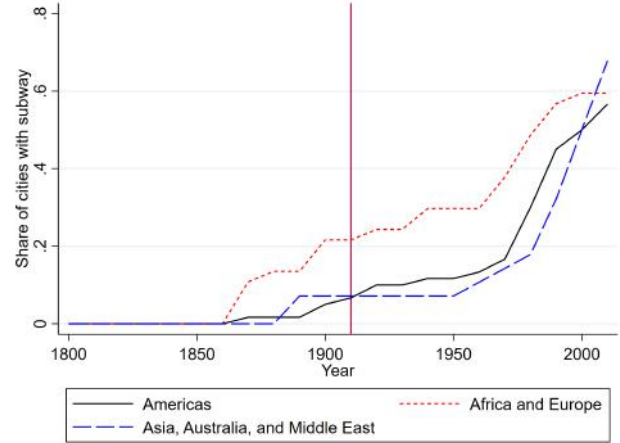
Last, Figure A9 summarizes the distribution of densities, wages, rents and commuting cost by distance from the CBD in the initial and final equilibrium. In keeping with the canonical model, rents decrease in distance from the CBD as commuting costs increase. Wages are generally higher in the CBD owing to exogenous and endogenous productivity advantages. Our shocks hit the largest concentrations of employment by design, with the consequence that there is an increase in dispersion of employment densities at close distances.

A.6 First nature, second nature, and transport revolution

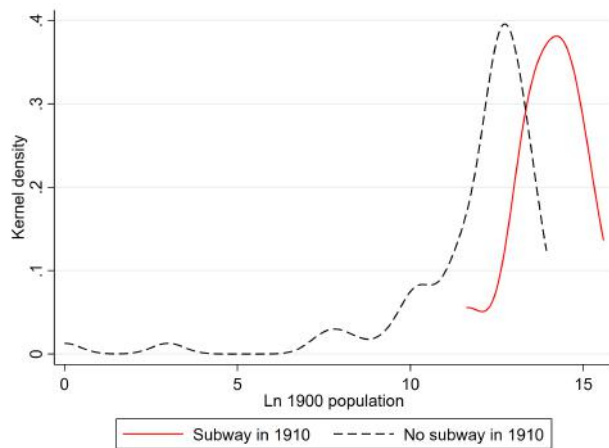
Figure A10: Transport and telecommunication through the ages



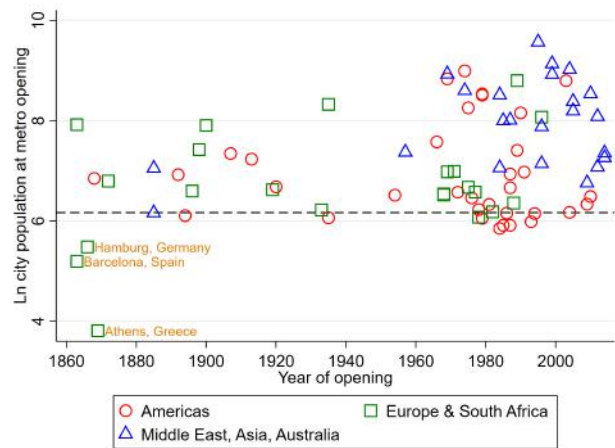
(a) Cars and Telephones per capita (United States)



(b) Subway adoption by world region



(c) Subway adoption by historic city size



(d) Metro openings and the population threshold

Notes: 4(a): For cars, we assume that there are 0 before 1900 (in 1900 there were 8,000). Sources: population and telephones are from Such and Carter, eds (2000), car data are from Federal Highway Administration Office of Highway Information Management (n.d.). 4(b-d): own calculations (see text) and Ahlfeldt et al. (2020b) for underlying data sources.

The Transport and Communication Revolution. Panel 4(a) of Figure A10 illustrates the extent of the (individual) transport and telecommunication revolutions throughout the 20th century.

Evolution of subway systems. Based on our 125-city sample, panel (b) of Figure A10 illustrates the adoption of the subway technology by opening years and world region. While European cities were early adopters, American and Asian cities have caught up over time.

Recently, Asian cities have even overtaken European ones in terms of the share of cities with a subway system. Panel (c) of Figure A10 shows that cities that were larger in 1900 were substantially more likely to have a subway system in 1910.

Minimum city size for rapid transport adoption. In the pre-automobile and pre-rapid urban transport age, the natural barrier to the extent of a city was closely related to and indeed limited by the ability to walk the city. See (Bairoch, 1988, p. 279) or (Daniels and Warnes, 2007, p. 2f). This allows us to calculate a population threshold, which results—when passed—in an increased demand in transportation as the city would have to expand to accommodate more citizens. Specifically, the population threshold is given by:

$$\text{POP}_{\text{threshold}} = \text{Area}_{\text{max}}^{\text{City}} \times \text{POPENSITY}_{\text{max}}^{\text{city}}, \quad \text{where} \quad \text{Area}_{\text{max}}^{\text{City}} = \pi \times r_{\text{max}}^2.$$

Based on data on population densities *before* the advent of rapid public transport, we can derive a plausible maximum population density $\text{POPENSITY}_{\text{max}}^{\text{city}}$ from historical data and then solve for the population threshold $\text{POP}_{\text{threshold}}$.³⁹

Table A 7: Population densities at the turn of the century

City	Population 1860 (1,000s)	Population 1900 (1,000s)	Area (km^2)	Density 1860	Density 1900
London	2,302	2,756	303	7,599	9,096
New York	814	3,437	57	14,275	60,302
Paris	1,696	2,714	78	21,744	34,795
Berlin	601	2,712	63.5	9,460	42,712
Average				13,269	36,726

Notes: Area is from Mattersdorf (1907, p. 11) and refers to city boundaries towards the end of the 19th century. We assume that for London, New York, and Paris, the boundaries changed as little between 1860 and 1900 as they did for Berlin.

Table A 7 reports 1860 and 1900 population density estimates for the four largest cities (as of 1900) in our sample. On average, the population density in the mid-19th century was about 13,500 people per square kilometer. However, the areal definitions of New York (only Manhattan) and Paris are more uncertain than the more clearly defined ones from Berlin and London. As data quality is best for Berlin, we take the value of 9,460 per square kilometer 1860.

In line with Bairoch (1988, p. 279) and as the typical maximum straight line distance that a human walks within one hour should be a bit smaller in the city than in a territory without obstacles, we assume $r_{\text{max}}^{\text{walking}} = 4\text{km}$. We assume a regular-circle shape of the city to calculate the resulting maximum area pre-transport: $\text{Area}_{\text{max}}^{\text{City}}$. Our resulting threshold is 475,511. When passing this threshold, the demand for transportation in a city would rapidly

³⁹Bairoch (1988, p. 279) assumes a maximum population density of 350 per hectare, but this seems too large of an average for whole cities (rather than the most dense ward).

increase as inhabitants would need to overcome distances by means other than walking. They would then request public transport.

Does this threshold correspond to the observed introductions of rapid urban transport systems? Panel 4(d) of Figure A10 plots the date of the metro openings in our sample against the threshold. Three outliers of the early days stand out. This, however, has to do rather with the coding than with the actual matter. In Barcelona and Hamburg, the start date is set to the respective dates as suburban or intra-urban railways were opened that would later be incorporated into the rapid transport system. The corresponding alternative opening dates of the metro system proper are 1924 for Barcelona (Busetti, 2015) and 1906 for Hamburg (Heinsohn, 2013, p. 39), dates by which each had long passed the threshold. In Athens, parts of the line that would become a part of a proper metro system was built in 1866 to connect the city with the port of Piraeus and electrified in 1904 (Mega, 2016, p. 135). If this date is taken as the metro opening rather than the extension of this line in 1926 (Pantoleon, 2010, p. 17), Athens remains an outlier. In spite of these outliers, the empirical evidence, just like historical narrative evidence, points to the existence of a threshold, which requires to be explicitly modelled.

Shocks in the first-nature world. Table A 8 shows the effects of disasters in the first nature era compared to those in the second nature era. The sample is restricted to cities that existed in 1800. The outcome variable in columns (1) and (2) is the distance between the city hall in 1900 and the first political institution—often dating back multiple centuries. It can be seen that early disasters have no significant effects on early city hall re-locations, but that disasters since 1900 have a strong effect on city hall re-locations between 1900 and 2000 (columns (3) and (4)).

Table A 8: The effects of city size and disasters on city hall relocations

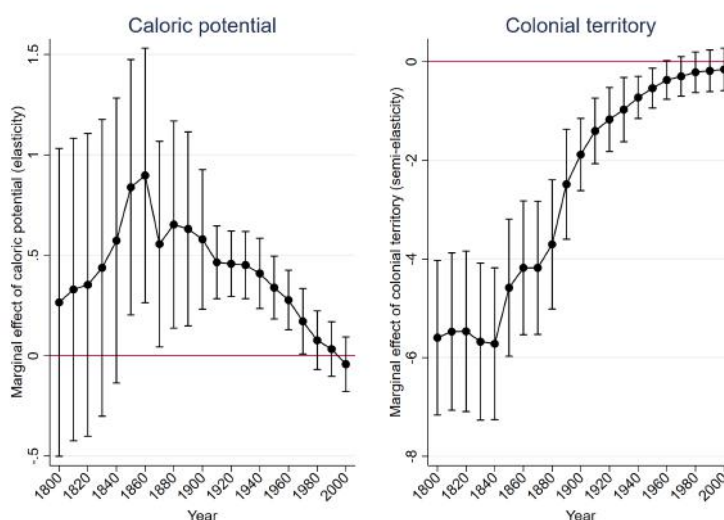
	(1)	(2)	(3)	(4)
	Ln distance between first and 1900 city hall	Distance between first city hall and 1900 city hall	Ln distance between 1900 and 2000 city hall	Distance between 1900 and 2000 city hall
Estimator	OLS	Poisson	OLS	Poisson
Ln population 1800	-0.039 (0.04)	0.226 (0.16)		
Disasters until 1900	-0.066 (0.06)	0.130 (0.14)		
Ln population 1900	0.032 (0.08)	-0.268 (0.23)	-0.132* (0.06)	-0.219 (0.17)
Disasters since 1900			0.133** (0.04)	0.241*** (0.07)
Ln pop. 1900	Yes	Yes	Yes	Yes
Ln pop. 2000			Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes
Observations	76	76	76	76
R^2 / Pseudo R^2	.0692	.0902	.432	.339

Notes: For details on the underlying data, see Global Cities dataset appendix (Ahlfeldt et al., 2020b). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A.7 Instruments for population

Figure A11 shows that our instrumental variables *caloric potential* and *colonial territory* are significantly correlated with city population in 1900, but not in 2000.

Figure A11: Effect of population IVs on population by year



Notes: Each dot represents a point estimate from year-specific regressions of ln population against ln caloric potential and the colonial territory dummy. For details on the construction of the instrumental variables and on the underlying data, see main text and Global Cities dataset appendix (Ahlfeldt et al., 2020b).

A.8 Empirical results - OLS estimates

Table A 9: The effects of 1900 population and disasters on spatial concentration - OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.139*** (0.02)	0.216*** (0.04)	0.249*** (0.04)	0.103*** (0.02)	0.167*** (0.03)	0.265*** (0.05)	0.247*** (0.07)
Disasters since 1900	-0.117*** (0.04)	-0.118** (0.06)	-0.201*** (0.06)	-0.133*** (0.03)	-0.122*** (0.04)	-0.168*** (0.06)	-0.049 (0.08)
Ln pop. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	125	125	125	125	125	125	125
R ²	.428	.383	.471	.283	.43	.379	.292

Notes: See Table 5.

Table A 10: Agglomeration- vs. transport-induced path dependency - OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.120*** (0.02)	0.198*** (0.04)	0.220*** (0.03)	0.086*** (0.02)	0.146*** (0.02)	0.246*** (0.05)	0.234*** (0.07)
Subway in 1910	0.351** (0.15)	0.353 ⁺ (0.23)	0.494* (0.28)	0.447*** (0.13)	0.407** (0.16)	0.414 (0.30)	0.362 (0.32)
Disasters since 1900	-0.108*** (0.04)	-0.109* (0.06)	-0.182*** (0.06)	-0.128*** (0.04)	-0.112*** (0.04)	-0.159** (0.06)	-0.046 (0.09)
Ln pop. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subway 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	125	125	125	125	125	125	125
R ²	.449	.393	.491	.335	.452	.389	.295

Notes: See Table 5.

A.9 Robustness of multiple equilibria

Single instrument. Table A 11 shows results if only caloric potential is used as an instrument (instead of caloric potential *and* colonial territory).

Table A 11: Single IV - Caloric Potential only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.120* (0.06)	0.282*** (0.09)	0.269*** (0.10)	0.216*** (0.06)	0.181** (0.08)	0.386*** (0.14)	0.389** (0.16)
Disasters since 1900	-0.116*** (0.03)	-0.123** (0.05)	-0.202*** (0.06)	-0.141*** (0.04)	-0.123*** (0.04)	-0.176*** (0.06)	-0.059 (0.08)
Ln pop. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. IV	CP	CP	CP	CP	CP	CP	CP
KP-F (p-val.)	.035	.035	.035	.035	.035	.035	.035
Observations	125	125	125	125	125	125	125

Notes: See Table 5.

Disaster effects. Table A 12 shows the disaster coefficient estimated for the 5 different disaster combinations for the 7 concentration measures (35 IV regressions in the same specification as in Table 5). To keep the coefficients' magnitudes comparable, we instrument all disasters with the count of disasters conditional on the respective restriction.

Table A 12: Robustness of disaster effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Disaster effect...							
...excluding fires	-0.092* (0.05)	-0.069 (0.08)	-0.156* (0.08)	-0.143*** (0.05)	-0.107** (0.05)	-0.137* (0.08)	-0.050 (0.09)
...excluding recent	-0.145*** (0.04)	-0.171*** (0.06)	-0.215*** (0.06)	-0.130*** (0.04)	-0.151*** (0.04)	-0.189*** (0.06)	-0.069 (0.08)
...only man-made	-0.155*** (0.05)	-0.223*** (0.08)	-0.213*** (0.07)	-0.152*** (0.04)	-0.146*** (0.05)	-0.216*** (0.07)	-0.061 (0.09)
...only natural	-0.096* (0.05)	-0.036 (0.07)	-0.204** (0.09)	-0.128** (0.05)	-0.115** (0.05)	-0.132 ⁺ (0.08)	-0.038 (0.11)
...including all	-0.122*** (0.04)	-0.125** (0.05)	-0.206*** (0.06)	-0.140*** (0.04)	-0.128*** (0.04)	-0.173*** (0.06)	-0.052 (0.08)

Notes: See Table 5.

Sources of variation: within- and between-country variation. In Table A 13, we add country fixed effects to our baseline specification to restrict the identifying variation to that origination from within countries. Results remain qualitatively similar and significance levels stay high for the population variable. The precision is reduced for the disaster variable, likely owing to the rare occurrence of disasters and the reduced sample size.

Table A 13: Sources of variation: within-country

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.082*** (0.02)	0.243*** (0.04)	0.131*** (0.04)	0.252*** (0.04)	0.098*** (0.03)	0.238*** (0.03)	0.237*** (0.03)
Disasters since 1900	-0.056** (0.03)	-0.047 (0.04)	-0.154*** (0.03)	-0.057 (0.04)	-0.055** (0.03)	-0.069 (0.05)	0.008 (0.05)
Ln pop. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. IV	CP & CT	CP & CT	CP & CT	CP & CT	CP & CT	CP & CT	CP & CT
KP-F (p-val.)	.304	.304	.304	.304	.304	.304	.304
Hansen J (p-val.)	.139	.156	.22	.276	.139	.136	.059
Observations	108	108	108	108	108	108	108

Notes: See Table 5.

In Table A 14, we use the mean values of historic city size and the number of disasters across cities within countries as instrumental variables to restrict the identifying variation to that originating from between countries. For most specifications, the results are statistically significant, qualitatively identical, and quantitatively close to our main results. If anything, the disaster effects are larger (though less precisely estimated).

Table A 14: Sources of variation: between-country

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.151*** (0.04)	0.202*** (0.05)	0.319*** (0.08)	0.103*** (0.03)	0.182*** (0.04)	0.303*** (0.08)	0.259*** (0.09)
Disasters since 1900	-0.192 (0.15)	-0.215 (0.19)	-0.329* (0.18)	-0.353*** (0.11)	-0.221 ⁺ (0.15)	-0.380** (0.18)	-0.235 ⁺ (0.16)
Ln pop. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. & disaster IV	Country means	Country means	Country means	Country means	Country means	Country means	Country means
KP-F (p-val.)	.001	.001	.001	.001	.001	.001	.001
Observations	125	125	125	125	125	125	125

Notes: See Table 5.

Sources of variation: North-America vs. rest of the world To allow for heterogeneity by world region, we interact our key variables of interest with two dummy variables that either indicate North American cities $I(NA = 1)$ or the rest of the world $I(NA = 0)$, allowing for a North-America-specific intercept. As Table A 15 shows, the point estimates are qualitatively the same in both parts of the world.

Table A 15: Sources of variation: between-country

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
NA=0 × Ln population 1900	0.211*** (0.08)	0.269** (0.12)	0.474*** (0.14)	0.100 (0.08)	0.302*** (0.09)	0.407** (0.17)	0.327* (0.17)
NA=1 × Ln population 1900	0.085** (0.04)	0.205*** (0.08)	0.145** (0.06)	0.207*** (0.08)	0.101** (0.05)	0.197** (0.10)	0.220* (0.12)
NA=0 × Disasters since 1900	-0.079** (0.04)	-0.091 ⁺ (0.06)	-0.181** (0.07)	-0.171*** (0.03)	-0.089** (0.04)	-0.175*** (0.06)	-0.077 (0.07)
NA=1 × Disasters since 1900	-0.104*** (0.04)	-0.086 (0.08)	-0.186** (0.08)	-0.037 (0.07)	-0.113** (0.05)	-0.091 (0.11)	0.011 (0.17)
Ln pop. 2000 × NA	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. IV	CP×NA & CT×NA	CP×NA & CT×NA	CP×NA & CT×NA	CP×NA & CT×NA	CP×NA & CT×NA	CP×NA & CT×NA	CP×NA & CT×NA
KP-F (p-val.)	.003	.003	.003	.003	.003	.003	.003
Hansen J (p-val.)	.512	.323	.86	.106	.419	.282	.278
Observations	125	125	125	125	125	125	125

Notes: See Table 5 for general notes. CP×NA & CT×NA is the interaction of caloric potential and North America (NA) effects as well as the interaction of colonial territories and NA effects. Similarly, ln pop. 2000 × NA is the interaction of ln 2000 population and NA effects (including a non-interacted NA effect). Controls include: Developed area 2000, irregular shape index, fragmentation index, share of land not developable within 5km of 1900 city hall, distance from 1900 city hall to water, share of land with steep slope. For details on the construction of IVs, see Global Cities dataset appendix (Ahlfeldt et al., 2020b).

A.10 Zero Stage

As outlined in the main text, it is likely that in predicting historic metro systems, subway potential interacts with historic city size non-linearly. To account for interactions between the exogenous shifters of historic population and subway adoption whose functional form is a priori unknown, we estimate the following zero-stage models: $\ln P_i^1 = \mathcal{P}(CP_i, SP_i, CT_i) + \varepsilon_i^{\mathcal{P}}$ and $M_i^1 = \mathcal{M}(CP_i, SP_i, CT_i) + \varepsilon_i^{\mathcal{M}}$, where CP_i is the caloric potential, CT is the dummy for being in colonial territories in 1800, and SP_i is the subway potential. \mathcal{P} and \mathcal{M} are functions that allow for flexible interactions between the exogenous population and subway shifters. M_i^1 is the subway indicator. $\varepsilon_i^{\mathcal{P}}$ and $\varepsilon_i^{\mathcal{M}}$ are residual terms. To estimate the functions \mathcal{P} and \mathcal{M} , we employ LWR regressions of the following type: $\mathcal{Y}_i = c_i^{\mathcal{Y}} + b_i^{\mathcal{Y}} I(CT_i = 1) + \varepsilon_{i,l}^{\mathcal{Y}}$, where $\mathcal{Y} \in \{\ln P_i^1, M_i^1\}$, $I(\cdot)$ is an indicator function returning one if the condition is met and zero otherwise, and $c_i^{\mathcal{Y}}$ and $b_i^{\mathcal{Y}}$ are city-specific parameters to be estimated. In each city-specific LWR $l \in N$, we weight observations i by a multiplicative Gaussian kernel:

$$W_{i,l} = \frac{w_{i,l}}{\sum_i w_{i,l}}, \quad w_{i,l} = \prod_{\mathcal{R}} w_{i,l}^{\mathcal{R}}, \quad w_{i,l}^{\mathcal{R}} = \frac{1}{\kappa^{\mathcal{R}} \sqrt{\pi}} \exp \left[-\frac{1}{2} \left(\frac{\mathcal{R}_i - \mathcal{R}_l}{\kappa^{\mathcal{R}}} \right)^2 \right],$$

where $\mathcal{R} \in \{CP, SP\}$ and κ^{CP} and κ^{SP} are bandwidth parameters to be chosen. In the baseline specification, we quite aggressively select narrow bandwidths of $\kappa^{CP} = 0.0075$ and $\kappa^{SP} = 0.1$ to maximize the strength of the instrument. As a robustness check, we employ rule-of-thumb values that are similar for the subway potential but about 10 times larger for the caloric potential.

From the LWR-estimates it is straightforward to construct two instrumental variables to identify historic population and subway effects: $IV_i^P = \ln P_i^1 - \varepsilon_{i=l}^{\mathcal{P}}$, $IV_i^M = M_i^1 - \varepsilon_{i=l}^{\mathcal{M}}$. Intuitively, this approach allows for a three-way interaction between CP_i , SP_i , and CT_i by allowing both the intercept $c_i^{\mathcal{Y}}$ and the marginal effect $b_i^{\mathcal{Y}}$ of CT_i to vary non-linearly in CP_i and SP_i due to the local weighting. By construction, both instrumental variables are exclusively based on data that are external to our 125 cities.

In Table A 16, we illustrate the explanatory power of the two instrumental variables that are generated in the zero stage regressions. Evidently, the population instrument from the zero stage is a strong (and exclusive) predictor of historic population (column 1). Likewise, the subways instrument strongly predicts the propensity of having a historic subway system (column 4). Hence, the two instrumental variables have the potential to separately identify historic population and subway effects. They are also relevant. The degree of explanatory power becomes evident as the R^2 almost doubles for the population in 1900 and quadruples for the adoption of the subway when we add the instruments on top of our set of geographic controls (columns (2) vs. (3) and (5) vs. (6)).

The population instruments should predict historic population monotonically and his-

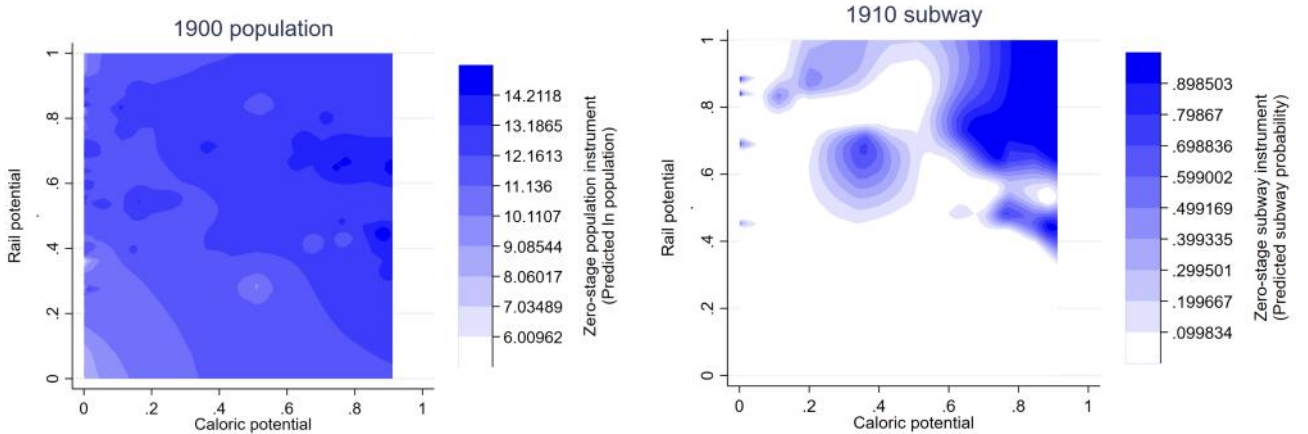
Table A 16: Relevance of zero-stage IVs

	(1)	(2)	(3)	(4)	(5)	(6)
	Log city population 1900	Log city population 1900	Log city population 1900	Subway in 1910	Subway in 1910	Subway in 1910
Zero-stage population IV	1.060*** (0.18)		0.965*** (0.19)	-0.010** (0.00)		
Zero-stage subway IV	0.067 (0.53)			1.227*** (0.07)		1.234*** (0.07)
Controls	-	Yes	Yes	-	Yes	Yes
Observations	125	125	125	125	125	125
R^2	.712	.399	.762	.843	.203	.868

Notes: Controls include Ln market access, disasters since 1900, ln 2000 population, 2000 subway system (dummy), developed area 2000, irregular shape index, fragmentation index, share of land not developable within 5km of 1900 city hall, distance from 1900 city hall to water, share of land with steep slope. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

toric subways only through an interaction with subway potential. Figure A12 illustrates how the exogenous historic population and subway shifters enter the zero-stage instruments.

Figure A12: Determinants of zero-stage instruments: LWR, narrow bandwidth



Notes: To generate the figure, we separately predict ln 1900 population and subway adoption using the locally weighted regressions with kernel weights based on caloric and subway potential (see text for details).

In keeping with our expectations, there is a strong interaction between caloric potential and subway potential in the subway instrument, whereas the relationship between the population instrument and caloric potential is more monotonic. Hence, we are confident that the two instruments separately identify the channels we are interested in.

A.11 Robustness Multiple Equilibria

Rule-of-thumb bandwidth. In the models reported in Table 6 we use very narrow bandwidths in the LWR-zero-stage to maximize predictive power and the efficiency in the second stage. As a robustness check, we replicate the analysis using the Silverman (1986) rule-of-thumb bandwidths of $\kappa^{\mathcal{R}} = 1.06\sigma_{\mathcal{R}}N^{\mathcal{R}-\frac{1}{5}}$, which results in $\kappa^{CP} = 0.0099$ and $\kappa^{SP} = 0.12$.

To accommodate the non-linear interaction between the determinants of subway systems with substantially increased bandwidths, we allow for linear interactions between CP_i , SP_i , and $I(CT_i)$: $\mathcal{Y}_i = \sum_{u=0}^B \sum_{v=0}^B \sum_{w=0}^1 b_i^{\mathcal{Y},u,v,w} \left((CP_i)^u \times (SP_i)^v \times I(CT_i = w) + \varepsilon_i^{\mathcal{Y}} \right)$, where we set the polynomial order to $B = 1$. Table A 17 reports the corresponding results.

Table A 17: Robustness - Rule-of-thumb bandwidth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.085* (0.05)	0.235*** (0.07)	0.269*** (0.07)	0.112*** (0.04)	0.147*** (0.05)	0.390*** (0.08)	0.431*** (0.10)
Subway in 1910	0.758*** (0.20)	0.489 (0.34)	1.090** (0.45)	0.649*** (0.17)	0.717*** (0.22)	0.535 (0.43)	-0.027 (0.52)
Disasters since 1900	-0.098*** (0.04)	-0.127** (0.06)	-0.196*** (0.07)	-0.131*** (0.04)	-0.113*** (0.04)	-0.200*** (0.07)	-0.106 (0.08)
Ln pop. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subway 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market access	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. & subway IV	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage
Kleinb.-Paap F (p-val.)	0	0	0	0	0	0	0
Observations	125	125	125	125	125	125	125

Notes: See Table 6.

Parametric Zero Stage. In a further alteration, we implement a fully parametric zero stage. To maintain a high degree of flexibility in the absence of a local weighting, we increase the polynomial order to $B = 2$.

$$\ln P_i^1 = \sum_{u=0}^B \sum_{v=0}^B \sum_{w=0}^1 b_{u,v,w} \left((CP_i)^u \times (SP_i)^v \times I(CT_i = w) \right) + \theta^P \ln(P^P) + \delta^P M_i^2 + X_i b^P + \varepsilon_i^P$$

$$\ln M_i^1 = \sum_{u=0}^B \sum_{v=0}^B \sum_{w=0}^1 b_{u,v,w} \left((CP_i)^u \times (SP_i)^v \times I(CT_i = w) \right) + \theta^M \ln(P^M) + \delta^M M_i^2 + X_i b^M + \varepsilon_i^M$$

Correspondingly, the two instrumental variables are:

$$IV_i^P = \sum_{u=0}^B \sum_{v=0}^B \sum_{w=0}^1 \hat{P}_{u,v,w} \left((CP_i)^u \times (SP_i)^v \times I(CT_i = w) \right)$$

$$IV_i^M = \sum_{u=0}^B \sum_{v=0}^B \sum_{w=0}^1 \hat{M}_{u,v,w} \left((CP_i)^u \times (SP_i)^v \times I(CT_i = w) \right)$$

Conditional on controls (also in the first and second stages), both instrumental variables are exclusively based on data that are external to our 125 cities. Table A 18 reports the results from the parametric zero stage.

Table A 18: Robustness - Parametric zero stage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Ln distance between EWPPs x (-1)	Ln CD of bilateral EWPP distances at 0.75 km	Ln average bilateral distance between PLs (x-1)	Distance from 1900 CH gradient (x-1)	Ln dist. from EWPP to 1900 CH (x-1)	Ln mean dist. from all PLs to 1900 CH (x-1)	Ln mean dist. from global PLs to 1900 CH (x-1)
Ln population 1900	0.150** (0.06)	0.303*** (0.09)	0.261*** (0.09)	0.155*** (0.05)	0.226*** (0.07)	0.462*** (0.10)	0.519*** (0.13)
Subway in 1910	0.932*** (0.34)	0.740 (0.54)	1.658** (0.69)	0.633** (0.30)	0.819** (0.35)	0.495 (0.59)	-0.158 (0.69)
Disasters since 1900	-0.107*** (0.04)	-0.135** (0.06)	-0.190*** (0.07)	-0.138*** (0.04)	-0.125*** (0.04)	-0.212*** (0.07)	-0.122 ⁺ (0.08)
Ln pop. 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subway 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market access	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1900 pop. & subway IV	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage	Zero stage
Kleinb.-Paap F (p-val.)	0	0	0	0	0	0	0
Observations	125	125	125	125	125	125	125

Notes: See Table 6.

Appendix References

- Ahlfeldt, Gabriel M. and Daniel P. McMillen**, "Tall buildings and land values: Height and construction cost elasticities in Chicago, 1870-2010," *The Review of Economics and Statistics*, 2018, 100 (5), 861–875.
- Bureau of Economic Analysis**, "Interactive access to industry economic accounts data: GDP by industry (Release April 19, 2018)," Technical Report, Bureau of Economic Analysis 2018.
- Buseti, Simone**, *Governing metropolitan transport: Institutional solutions for policy problems*, Springer International Publishing, 2015.
- Combes, Pierre-Philippe and Laurent Gobillon**, "The empirics of agglomeration economies," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 247–348.
- Federal Highway Administration Office of Highway Information Management**, "Highway statistics summary to 1995," Technical Report.
- Heinsohn, Ralf**, *Schnellbahnen in Hamburg: Die Geschichte von S-Bahn und U-Bahn, 1907-2007*, BoD, 2013.
- Mattersdorf, Wilhelm**, *Städtische Verkehrsfragen*, Springer. Berlin, 1907.
- Mega, Voula P**, *Conscious coastal cities*, Springer, 2016.
- Pantoleon, Skayannis**, "Project profile–Greece–Athens Metro," Technical Report, UCL Barlett School of Planning 2010.
- Silverman, Bernard W.**, *Density estimation for statistics and data analysis*, Chapman and ed. 1986.
- Such, Richard and Susan B. Carter, eds**, *Historical statistics of the United States—millennial edition online*, Cambridge University Press, 2000.