Prime locations*

Gabriel M. Ahlfeldt† Thilo N. H. Albers‡ Kristian Behrens§

October 28, 2021

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. Prime services are less spatially concentrated and prime locations are farther away from historic cores in historically smaller cities that did not develop early public transit networks. We rationalize these novel stylized facts empirically and theoretically. External returns to scale give rise to multiple equilibria in the city-internal distribution of prime services. The resilience of historic prime locations in historically large cities originates at least partially from endogenous durable transport networks.

Keywords. Prime services; internal city structure; transport networks; multiple equilibria and path dependence; simulation model.

JEL classification. R38, R52, R58.

*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 Abu Dhabi (NYU, online), Barcelona (Autonoma), London (SERC), Düsseldorf (EGIT), Toronto (Sauder-UBC workshop), Madrid (CUNEF, online), Montreal (UQAM-Bank of Canada workshop), Saint Petersburg (HSE), the Urban Economics Association Virtual Meeting, the Online Spatial and Urban Seminar, and especially Nate Baum-Snow, Victor Couture, 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; SSHRC Insight Grant # 435-2016-1246; and the HSE University Basic Research Program. Albers gratefully acknowledges the support of the CRC TRR 190 (DFG). The usual disclaimer applies.

†London School of Economics and Political Sciences (LSE), CEPR, CESifo, and CEP; g.ahlfeldt@lse.ac.uk.

‡Humboldt University Berlin, and Lund University; thilo.nils.hendrik.albers@hu-berlin.de.

§Université du Québec à Montréal (UQAM), HSE University, and CEPR; behrens.kristian@uqam.ca.

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. Prime locations historically emerged at places with favorable geographic features such as navigable rivers. These historic prime locations proved more resilient over the 20th century in cities that were historically dense and had developed early public transport networks. Second, we provide causal evidence that rationalizes heterogeneity in the spatial configuration of prime services across global cities. Major disasters over the 20th century lead to more dispersion in the spatial distribution of prime services which we measure in a variety of ways. Using instruments for 1900 population—exploiting soil conditions and colonial history—we show that historic city size anchors urban structure and leads to more spatially concentrated prime services and contemporary prime locations that are closer to historic city cores. Third, we shed light on the theoretical mechanisms that underlie these reduced-form relationships. Empirically, we show that disasters before 1900 had no significant effect on historic prime locations, suggesting a role for external returns to scale. Using a novel instrument for early subway adoption—subway potential—we show that the anchoring effect originates, at least partially, from durable transport networks that form in larger cities. We substantiate these empirical insights in a series of Monte Carlo experiments where we feed a granular version of a quantitative spatial model into the data generating process. We can reproduce the patterns in our data only when there are external returns to scale in the prime services sector and we allow for endogenous transport network formation.

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 is especially acute for prime services. Consequently, despite being more important than manufacturing for cities, they are less studied. 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 open-source 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 service establishments and the workers they employ. In particular, data on the construction of tall buildings around that period, the historic locations of bank and insurance offices, and the spatial distribution of prime service workers all suggest that historic prime service employment declined rapidly with distance 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 persistence in the spatial distribution

[1] Duranton 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.

of economic activity within cities.[2] Second, we document that there is less persistence 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).[3] 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 novel features that make historic prime locations more resilient: early densification and early mass transit systems, both of which prevail in historically large cities.

Guided by these stylized facts, we then provide the first empirical test for multiple spatial equilibria *within* cities and differences in their degree of inertia. Capitalizing on our big data approach, we compute several spatial concentration measures that capture the city internal spatial distribution of prime services and prime locations in 125 global cities. We combine these spatial concentration measures with major historic disasters and historic city size 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.

Finally, we explore the mechanisms that underlie the disaster and historic city size effects. 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. To disentangle transport-induced persistence from other

[2]The 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).

[3]A 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 (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.

channels, we use and augment data on historic subway systems from Gonzalez-Navarro and Turner (2018).[4] 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 instrument, combined with our previous population instruments, we find that the early development of mass transit systems has a large causal effect on the evolution of urban spatial structure 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%. 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 the historic city cores.

While we push the frontier of micro-geographic empirical research on long-run changes in internal city structure, natural experiments that switch off external returns and deploy mass transit networks do not exist. In the last step of our analysis, we therefore take control of the data generating process using simulation techniques. More precisely, we conduct a series of Monte Carlo experiments within a granular version of a quantitative spatial model that we develop in a companion paper (Ahlfeldt et al., 2021). The model features external returns that generate agglomeration of jobs, commuting costs that lead to co-location of jobs and workers, and a dispersion force that emerges from a less than perfectly elastic supply of real estate (Ahlfeldt et al., 2015).[5] We augment the model with endogenous transport network formation subject to increasing returns to scale (Fajgelbaum and Schaal, 2020; Santamaria, 2020). The distinctive element of the model is that indivisible workers

[4]Establishing 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.

[5]See Ahlfeldt and Pietrostefani (2019) and Duranton and Puga (2020) for reviews of the urban economics literature on spatial agglomeration and dispersion forces.

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). We distinguish between single-establishment prime services firms, global multi-establishment prime services firms, and non-prime services firms that compete for productive locations to generate one-for-one the same spatial concentration measures as in the empirical part of the paper using model-based equivalents of prime service establishments.

In each Monte Carlo experiment, we randomize city size as well as the assignment of productivity to workers and firms and solve for an equilibrium in the absence of external returns to obtain the model-based equivalent of a city observed in 1900 in our 'Gobal Cities' data set. The city may then develop a mass transit system before being eventually hit by a virtual disaster that forces a random number of firms to relocate away from the historic prime location. In a temporary equilibrium, only the shocked firms find optimal locations. In the final equilibrium, the model-based equivalent of a city observed in 2000, there is full re-optimization of firm-location choices and worker-firm assignments under external returns such that the spatial distribution of prime services firms may or may not revert back to the initial equilibrium. Running such Monte Carlo experiments several thousand times, we obtain a synthetic 'Global Cities' data set that we can subject to the same regression models as those used in the empirical parts of the paper. One advantage of the Monte Carlo setting is that we can generate exogenous and independent variation in city size and transport networks, hold the distribution of fundamentals constant, and observe the distribution of prime services in the historic period, which solves all identification problems. Moreover, we can surgically switch off external returns to scale and endogenous transport network formation. The takeaways are important. We can reproduce all patterns in the data solely through the interplay of exogenous shocks and economic forces that are canonical in recent urban models. External returns are a necessary condition for disasters to shift the spatial distribution of prime services between multiple equilibria and to lead to a more dispersed spatial configuration. The permanent locational advantage durable transport networks create around their central nodes represents a major source of inertia that provides a stronger anchoring effect than agglomeration economies do.

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 provides causal evidence on the role of historic city size, the development of early mass transportation networks, and disasters in shaping the spatial distribution of prime services within cities. Section 4 expores the underlying mechanisms empirically and theoretically (using Monte Carlo experiments). Section 5 concludes.

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 causal analysis and exploration of mechanisms in the remainder of the paper.

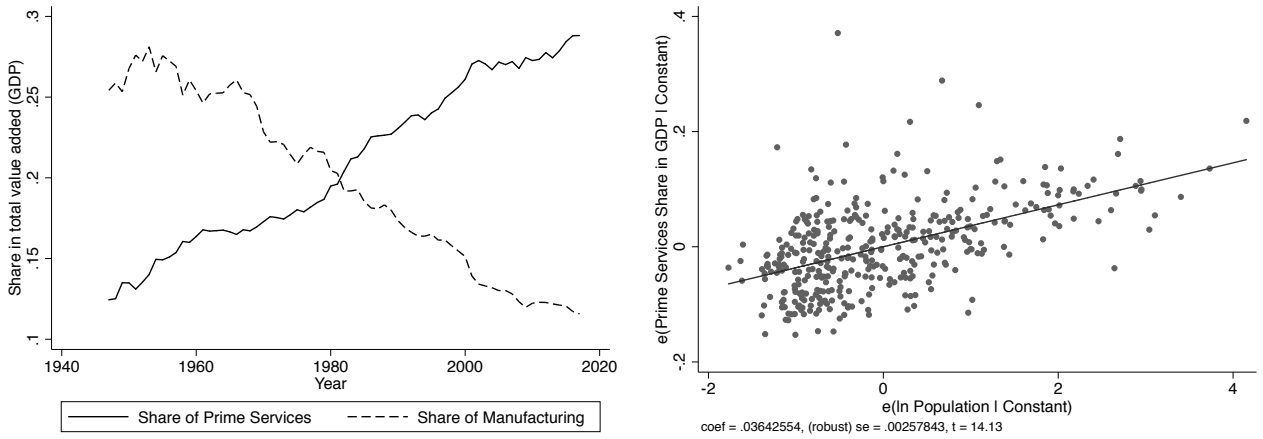
2.1 Geographic concentration of prime services

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. 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

[6]Recent 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



(a) Share of prime services in GDP

(b) Share of prime services and MSA size

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. (2020) for a description of the sources, details on the imputation, and results without imputation.

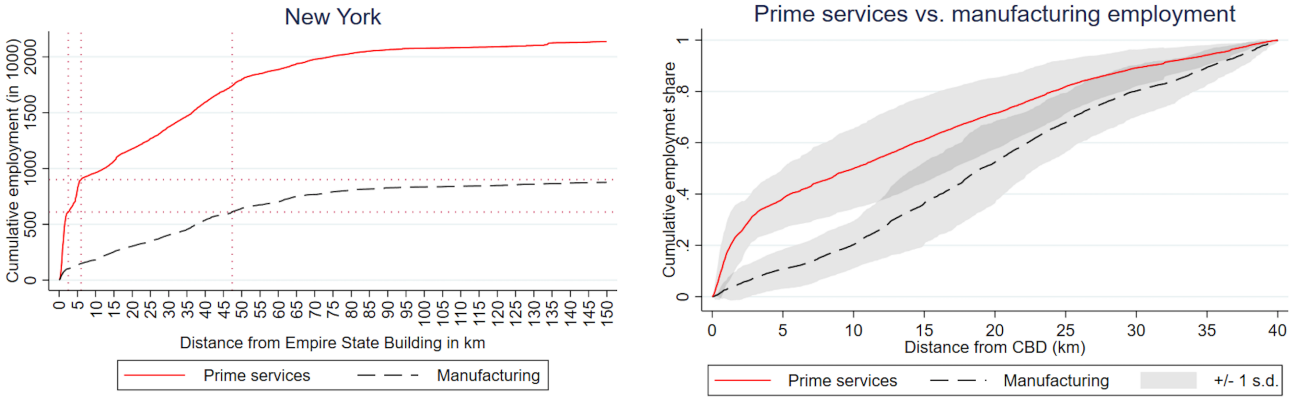
area, Champaign-Urbana, to the largest, New York-Newark-Jersey City, increases the prime service share by $[\ln(\text{pop}_{NY}) - \ln(\text{pop}_{CU})] \times \widehat{\beta} \approx 4.44 \times 0.036 = 16.28$ percentage points—a change roughly equivalent to moving the US economy from 1950 to 2015 (see 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

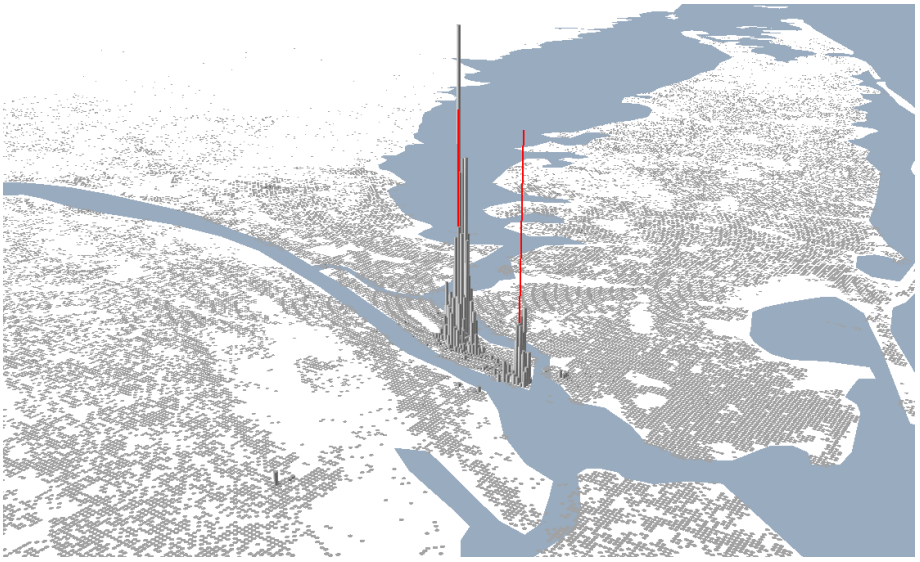
[7]See, e.g., Arzaghi and Henderson (2008), who study the clustering of a small subset of our prime service sector, advertising agencies, in Manhattan; and Lui et al. (2018), who provide evidence for within-building agglomeration economies for a subset of prime services (including law offices, advertising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



(b) Geographic concentration of prime service employment in NY

Notes: In Panel (a), the unit of analysis is city × 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 250×250m 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 250×250 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.

2.2 Predicting the location of prime services in cities

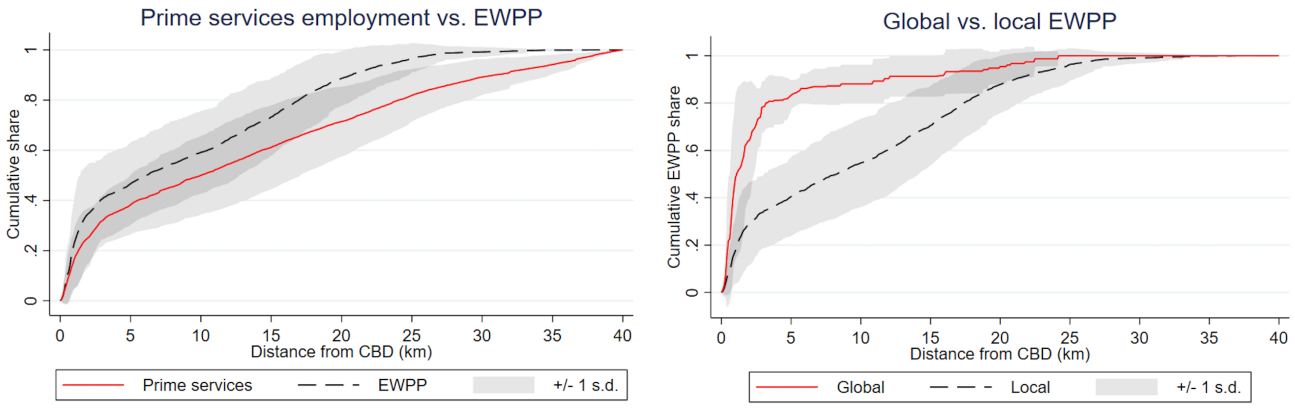
Ideally, we would use highly disaggregated micro-geographic data, such as those underlying Figure 2, to detect clusters of prime services. However, such data are only available for a limited number of cities, thus restricting substantially the global scope of the analysis. Following Henderson et al. (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 × 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 units 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 can 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. The predictive power is not limited to North American cities. In Table 1, we show

[8] At 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.

that our scraped PPs are good predictors of Starbucks franchises, grade-A office buildings held by real estate investment trusts (SNL-S&P investments), and co-working spaces in all continents, with North America being no particular outlier in terms of explanatory power. In Moscow, where we have access to geo-coded data for a very large sample of prime service establishments, we find that our PPs explain nearly two thirds of the variation in the number of establishments (see Figure A4).

Table 1: Predictive power of prime points by world region

Outcome	Stat.	Africa	Asia	Europe	N. A.[a]	Oceania	S. A.[b]
Starbucks	Coeff.	-	0.049***	0.029***	0.072***	0.008***	0.008
	R^2	-	0.211	0.399	0.284	0.616	0.188
SNL-S&P investments	Coeff.	0.194***	0.113*	0.070***	0.053***	0.072*	0.037***
	R^2	0.664	0.277	0.384	0.119	0.578	0.568
Coworking spaces	Coeff.	0.021**	0.040***	0.034***	0.039***	0.030**	0.025***
	R^2	0.350	0.241	0.519	0.350	0.787	0.278

Notes: [a]North America. [b]South America. Table reports regression coefficients and R^2 from regressions of a given outcome against prime points conditional on city fixed effects. Outcomes and prime points are measured as counts within 30,000 randomly drawn disks of 750-meter radii in each city. We discard disks with zero counts in both outcomes and prime points. In each column, we pool all global cities within a continent in one regression. * p < 0.1, ** p < 0.05, *** p < 0.01.

Last, the right panel of Figure 3 compares the concentration of companies with the largest global reach with that of all other prime service companies. 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 enables 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 some workplace amenities such as 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 cites, as recorded in the SNL-S&P Global data base, 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).

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 or at least 25 Starbucks franchises.

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 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

[9] We 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.

many edge cities (Lang, 2003, p. 65). We provide corresponding maps for each city in the Global Cities dataset Appendix (Ahlfeldt et al., 2020).[10]

Table 2: Data summary

Disaggregated data, address- or grid-level		**Urban biographies**
Economic data	Physical geography	Historical data
Core data	**Grid level**	**Address-level**
Central banks ⋆	Land cover	First settlement ⋆
Consultancy firms ⋆	Elevation	First political institution ⋆
Co-working spaces ⋆	Water cover	City hall 1900 ⋆
Insurance firms ⋆		City hall 2000 ⋆
Investment banks ⋆	**City level (built from grid)**	
Law firms ⋆	Developable land †	
Stock exchanges ⋆	Irregular shape index †	**City level**
	Fragmentation index †	Colonial occupation in 1800 ⋆
Employment weighting	Caloric potential of hinterland	Government type (1800-2000) ⋆
Scott's Business Directories (P)		Manmade disasters ⋆
NETS National Establishment		Natural disasters ⋆
Time Series (P)		Population (1800-2000) ⋆
		Rapid transit openings †
Validation data		Subway potential (1900) †
Geotagged photos		Market potential (1900) †
Geotagged Twitter tweets		
Co-working spaces ⋆		**Validation data**
Grade-A office buildings - SNL (P)		1880 US census (10% sample) †
Starbucks establishments		Member list of International
Emporis tall buildings (P)		Chamber of Commerce (1922) ⋆
US County Business Patterns		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., 2020).

Table 2 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

[10]Recent 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).

[11]These are improvements 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).

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., 2020). Finally, Table 2 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

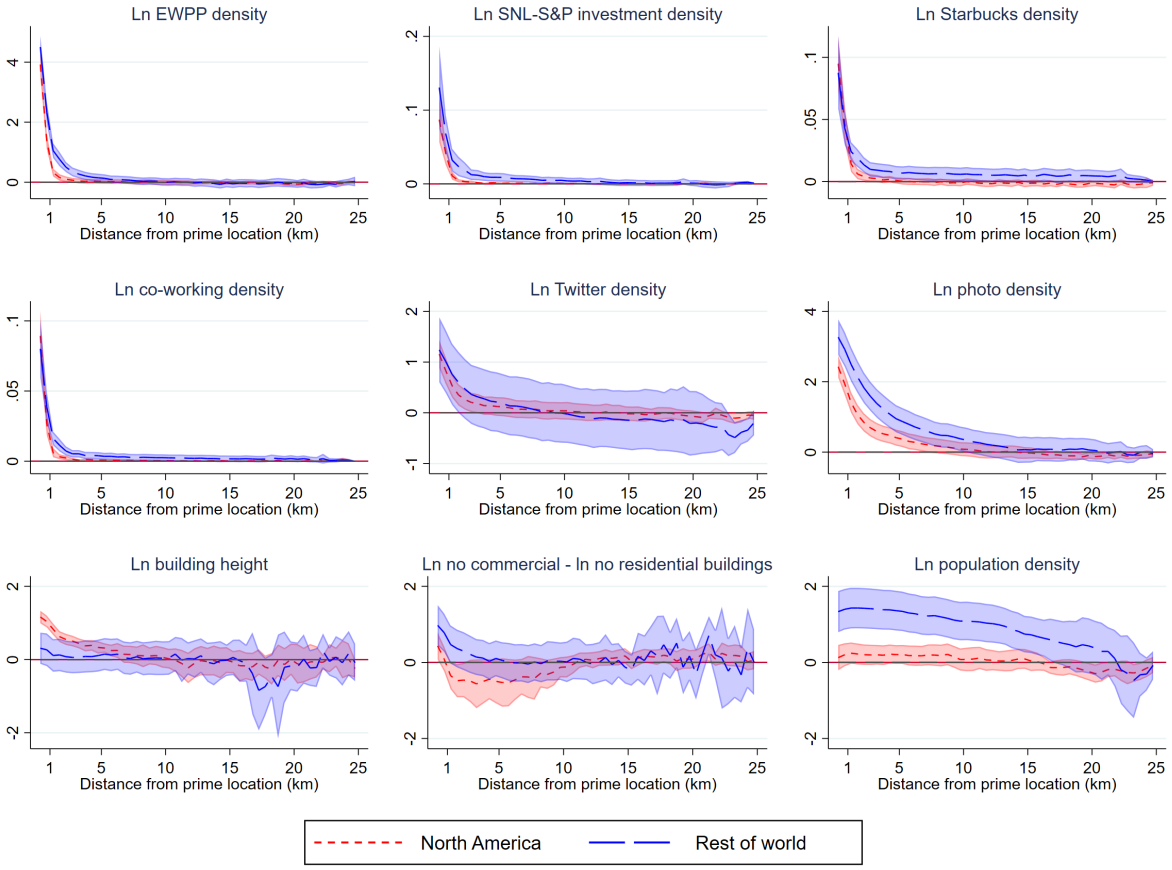
In our earlier discussion of New York and a handful of other North-American cities, we assumed the location of the central business district based (CBD) on what could be considered common sense (e.g., the Empire State Building for New York). Such an approach is neither satisfactory nor scalable to our 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). This makes it difficult to pin down a CBD.

For a more systematic identification of prime locations, 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

[12] In generating the 100k draws, we discard all draws that fall into undevelopable grid cells (see Ahlfeldt et al. (2020) 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.

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: 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., 2020).

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

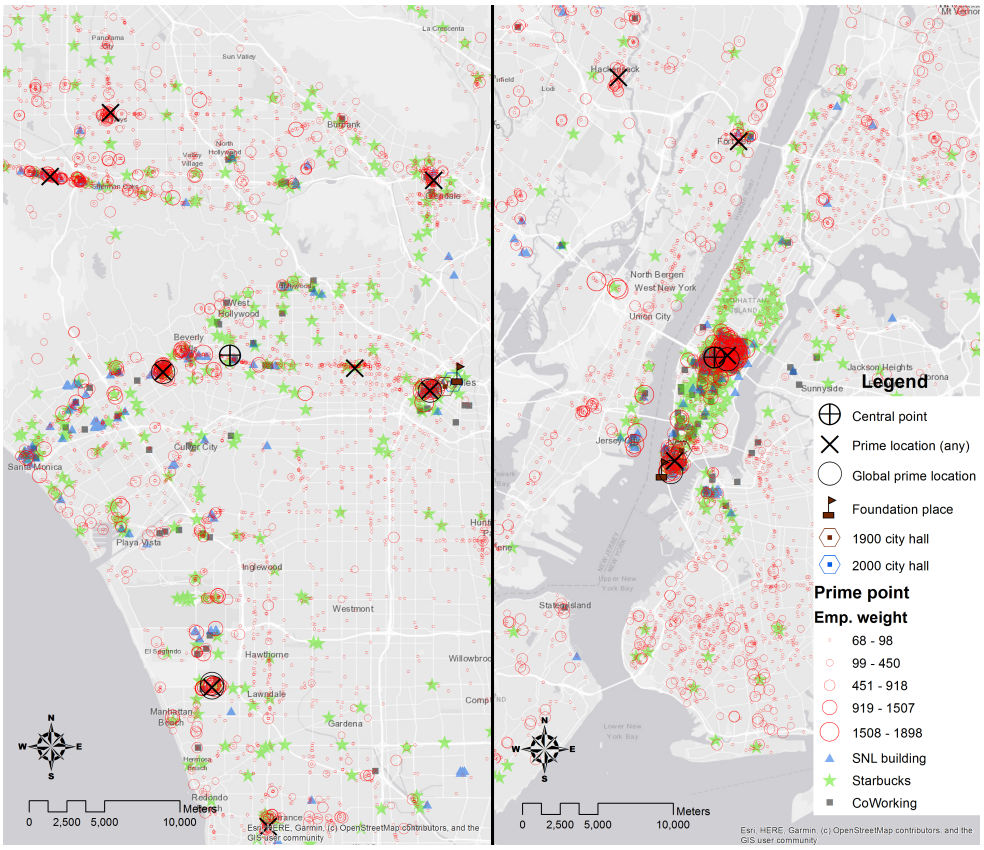
prime locations. This is reassuring on two fronts. First, co-working spaces are a relatively new phenomenon and they are arguably less prone to either tourism effects (Starbucks) or the legacy of long-term investments (SNL). Second, co-working spaces are more directly linked to the production of knowledge-based 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 with distance 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. We find that building height gradients are steeper in Canada and the USA than in the rest of the world, but population density gradients are flatter. Other than that the gradients are similar, lending further support to the external validity of the employment weights we assign to our prime points.

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).

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. Yet, the *central point*—intuitively denoting the location with the

[13] We 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: 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., 2020).

greatest accessibility to prime points within 5 km (see Appendix A.4)—has moved halfway from Downtown towards Santa Monica in LA, 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 geographies 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 lo-

cations 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. Do these insights generalize?

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.

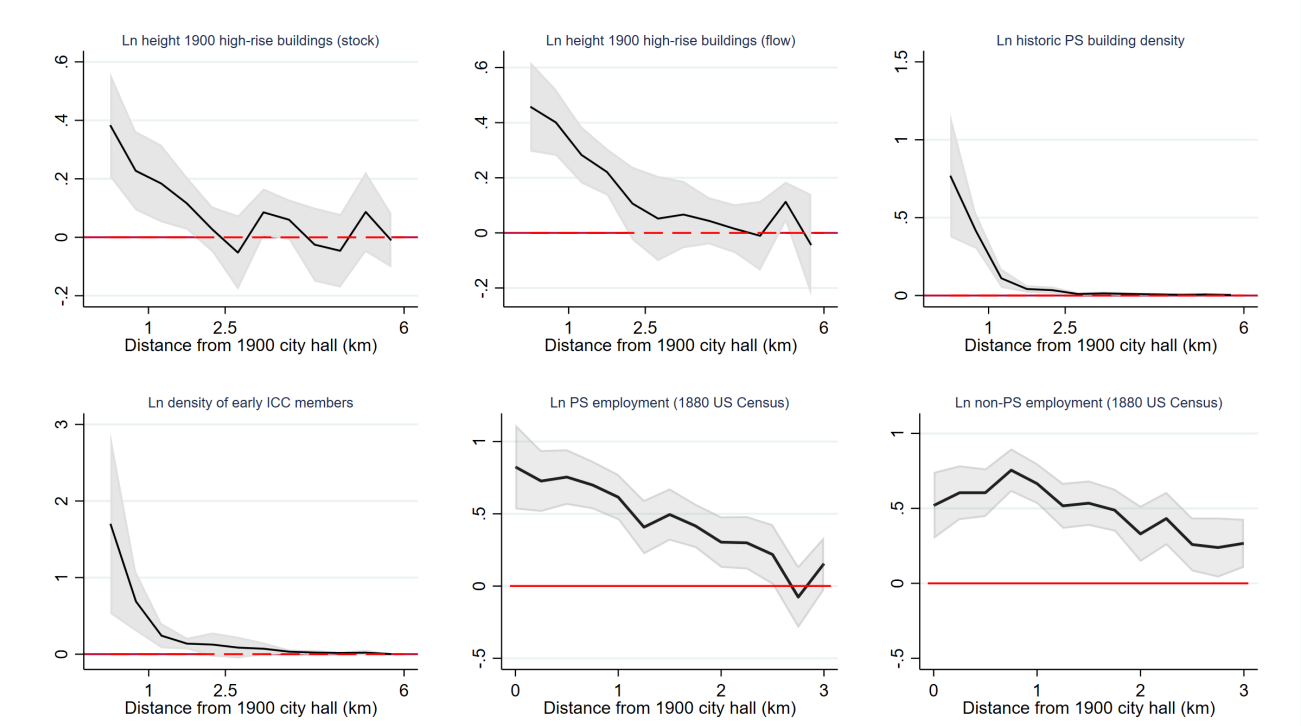
Canonical models of the economic geography literature consider the industrial era as the starting point for second nature attributes to matter. Beginning with this period, locational fundamentals become less important compared to external returns to scale that arise from input sharing, labour market 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 A5 in Appendix A.5). These reductions in transport and communication costs

[14] According 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.

induced urban areas to specialize according to their comparative advantage in interactive tasks (Michaels et al., 2018; Heblich 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 service activity.

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 (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., 2020).

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 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. Since 1880 predates the development of mass transit and the adoption of the automobile, the distribution of jobs should closely follow the distribution of people. It mimics the population gradient in Figure 5 in that it increases within the first kilometer as 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 persistence in global cities. Concretely, we are seeking to substantiate or reject the role of locational 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 investigation of causal mechanisms in Sections 3 and 4.

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 3 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

[15]Soil 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.

Table 3: 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***	1.320***	0.831***	0.558***	0.422**	0.681***	0.622+	0.549***
	(0.23)	(0.23)	(0.18)	(0.14)	(0.18)	(0.22)	(0.39)	(0.15)
Average elevation	-12.130**	-7.171***	-1.515+	-1.453	-1.706	-1.223	-5.774	-1.334
	(5.01)	(1.79)	(0.98)	(1.05)	(1.92)	(1.29)	(6.79)	(1.07)
Average slope	10.332**	2.243	-1.529	-4.207**	-1.037	-6.127***	9.640	-4.916***
	(5.21)	(2.30)	(1.70)	(1.73)	(3.42)	(1.91)	(7.79)	(1.78)
Bare ground	-1.860+	-1.368*	0.397	0.015	-10.714***	-0.041	.	0.001
	(1.18)	(0.77)	(0.70)	(0.55)	(0.79)	(0.57)	(.)	(0.55)
Foundation place in cell		5.599***						
		(0.24)						
Near 1900 city hall			4.098***	3.392***	3.730***	3.023***	3.634***	3.364***
			(0.12)	(0.09)	(0.13)	(0.14)	(0.27)	(0.10)
Cities	125	125	125	125	70	55	14	111
N (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 × 1.5km 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., 2020). + p<0.15, * p < 0.1, ** p < 0.05, *** p < 0.01.

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 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 3 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 3 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 persistence (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 3 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 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, our data suggest that cities that were larger in 1900 and were early adopters of mass transit remain more strongly centered on their historic cores nowadays. Hence, there appears to be a role for initial conditions—city size and early mass transportation networks—in moderating the occurence of multiple equilibria.

3 Causal evidence

Motivated by the stylized evidence in Section 2, we formulate two hypotheses that directly connect to theoretical research in urban economics and economic geography. We hypothesize that cities which experienced major spatial shocks such as natural disasters over the 20th century were more likely to undergo a transformation in the spatial distribution of prime services, away from the traditional monocentric pattern and towards a polycentric configuration with multiple prime locations farther away from the historic core. This hypothesis is consistent with the notion of multiple equilibria that prevails in urban models featuring either firm-level increasing returns (Fujita, 1988; Krugman, 1991) or external returns (Fujita

[16] In a first-nature world, once the shock is resorbed, economic activity should move back to its initial location (Davis and Weinstein, 2002). Table 5 in Appendix A.5 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.

[17] Early 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.

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

and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002; Ahlfeldt et al., 2015). We are the first to subject this intuition to an empirical test in a large set of global cities. We further hypothesize that historically larger and denser cities were more likely to remain monocentric. This hypothesis is consistent with a growing economic geography literature that explores the role of path dependency in spatial configurations theoretically (Allen and Donaldson, 2020) and empirically (Bleakley and Lin, 2012). Again, we are the first to engage with this phenomenon in a large sample of global cities. In this section, we provide reduced-form causal evidence supporting both hypotheses. We will turn to the underlying mechanisms, empirically and theoretically using simulations, in Section 4.

3.1 Identification

Separating the effect of second-nature geography from that of first-nature geography is a notorious challenge in economic geography and urban economics. In developing a transparent identification strategy, it is helpful to be explicit about the data generating process.

Let i denote cities and $t = 1, 2$ denote periods. 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. Assume there exists a *pure* measure of geographic concentration of prime services, $c_i^t = f(PSempl_i^t, \mathbf{u}_i^t)$, 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 $\mathbf{u}_i^t$. 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:

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

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. Through ϵ_i, we acknowledge that we are not able to observe all empirically relevant shocks.

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

$$C_i^t = c_i^t \times e^{a_i^t} \times (P_i^t)^{\theta^t}, \tag{2}$$

which includes c_i^t but also depends on locational fundamentals, a_i^t, and population, P_i^t. 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. Since there is an infinite number of ways to reduce city structure to a scalar, we consider several spatial concentration measures capturing different shades of the city-internal distribution of prime services (see Appendix A.4 for details): (i) the log of the average distance between EWPPs; (ii) the log of the CDF of distances between EWPPs evaluated 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; (vi) the log of the average distance of PLs to the 1900 CH; and (vii) the log of the average distance of global PLs to the 1900 CH.[19] We have no prior on the sign and magnitude of θ^t.

Substituting (2) into (1) and regrouping terms, we obtain:

$$C_i^2 = \alpha \times (C_i^1)^{\zeta} \times (P_i^1)^{\beta-\zeta\theta^1} \times (P_i^2)^{\theta^2} \times e^{\gamma D_i^2 + a_i^2 - \zeta a_i^1 + \epsilon_i} \tag{3}$$

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. It is impossible to approximate the spatial distribution of prime services in the historic period at the same spatial granularity as in the contemporary period. Therefore, we adopt the canonical assumption from neoclassical urban models that, in the absence of external returns, cities display a monocentric structure (Brueckner, 1987). This assumption implies that $C_i^1 = \bar{c}e^{\xi_i^1}$, where ξ_i^1 is an error term. From (3), we can then derive a reduced-form empirical specification 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:

$$\ln C_i^2 = \widetilde{\alpha} + \widetilde{\beta} \ln P_i^1 + \theta^2 \ln P_i^2 + \gamma D_i^2 + \widetilde{b} X_i + e_i, \tag{4}$$

with $\widetilde{\alpha} = \ln \alpha + \ln \bar{c}$, $\widetilde{\beta} \equiv \beta - \zeta\theta^1$, $\widetilde{b} \equiv b^2 - \zeta b_1$ and $e_i \equiv \epsilon_i + (\rho^2 - \zeta\rho^1) f_i + \zeta\xi^1$. Based on the stylized evidence in Section 2, we expect $\widetilde{\beta} > 0$, i.e. historically larger cities are more concentrated today. This will be the case if the centripetal force of historic population captured by β (e.g., originating from a more agglomerated and better connected historic center) exceeds the centrifugal force captured by $\zeta\theta^1$ (e.g., due to more likely emergence of sub-centres). Intuitively, we expect $\gamma < 0$ since displacement of prime services establishments due to temporary shocks can have permanent effects under external returns, leading to a more dispersed configuration.

An inspection of the error term reveals that unobserved location fundamentals affecting the internal structure of cities, f_i, will lead to a biased estimate of $\widetilde{\beta}$ if they are correlated with

[19] As 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 simulattion model in Section 4, we can observe the initial prime location and thus use it directly, contrary to the empirical exercise.

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 assume that our error in capturing shocks (ϵ_i) and historic effective concentration (ξ_i^1) are uncorrelated with our instruments IV_i^P and our shock measure D_i^2, and that the latter is uncorrelated with f_i.

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 unconducive 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.[20] 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

[20]The caloric potential instrument is arguably less likely to be endogenous than the colonial territories instrument. Our results are robust to using the caloric potential instrument only.

steam engine—the caloric potential of the hinterland was more important 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.6). This makes it less likely that they affect contemporaneous outcomes other than through persistence in historic city size.

Our measure for shocks since 1900, D_i^2, 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. Since we collect our data on major disasters from reliable urban biographies, there is little reason to expect measurement error (ϵ_i) to be correlated with historic population, its instruments, or the number of disasters. As for historic effective concentration (ξ_i^1), we can at least show that 1900 city hall gradients are uncorrelated with: (i) historic population; (ii) its instruments; and (iii) post-1900 disasters (see appendix Section A.6).

3.2 Results

Table 4: 1900 population and disaster effects on spatial concentration measures

	(1) Ln distance between EWPPs (×-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (×-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.213***	0.312***	0.333***	0.201***	0.255***	0.347***	0.284***
	(0.05)	(0.08)	(0.07)	(0.05)	(0.06)	(0.10)	(0.11)
Disasters since 1900	-0.122***	-0.125**	-0.206***	-0.140***	-0.128***	-0.173***	-0.052
	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)	(0.06)	(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., 2020) for sources and data construction. Robust standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Table 4 presents estimates of specification (4). In line with our empirical hypotheses, dis-

asters occurring after 1900 lead to cities with a more dispersed geography of prime services, whereas a larger 1900 city size anchors the spatial structure of the city, leading to more concentration around the historic core. By exploiting variation internal to global cities, these results provide a new perspective on the popular notion in economic geography that major shocks can shift a spatial economy to an alternative (more dispersed) steady state, whereas historically large agglomerations represent a source of inertia.

In column (1) of Table 4, we use the average bilateral distance between EWPPs as an intutitive 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.[21] 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 4 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 prime locations—hosting establishments of global leaders in the production of prime services—are the most resilient to shocks, consistent with initial

[21] The 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).

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 inertia (the population effect). This result survives a battery of robustness checks, with a natural starting point being the OLS estimates (see Appendix A.7). 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—in line with these shocks being 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.8 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: (i) fires, the likelihood and magnitude of which might be related to historic population density and thus not exogenous; (ii) natural disasters, or (iii) man-made disasters only; and (iv) recent disasters that might only cause a temporary deviation from a unique steady-state in contemporary cities (Davis and Weinstein, 2002). Finally, we analyze in more detail the source of our variation by identifying the effects separately from within- and between-country variation. Our results are not driven by North American cities only.

4 Mechanisms

So far, we have used our global cities dataset to establish that: (i) major disasters facilitate a transition from the traditional monocentric urban structure to more complex spatial configurations of prime services; and (ii) historic city size represents a source of inertia that works in the opposite direction. These findings help to rationalize the variety of city structures observed in the data and are broadly consistent with stylized predictions of economic geography and urban economics models that feature increasing returns to agglomeration, either internal or external to firms. The purpose of this section is to shed light on mechanisms that drive the reduced-form relationships. In particular, we wish to establish whether external returns were catalysts of urban transformation over the 20^{th} century and whether durable transport networks are a source of inertia as suggested by the stylized evidence in Section 2.

We proceed in two steps. In the first step, we expand on the empirical specification developed in Section 3. We add to the notion that external returns are key to multiple equilibria in the internal structure of cities by demonstrating that disasters had little effects on the location of urban cores during the first-nature geography era. Employing a novel instrument for early subway adoption, we show that durable transport networks account at least partially for inertia in the spatial structure of large cities. In the second step, we conduct a series

of Monte Carlo experiments on a synthetic data set generated by a granular spatial model developed in a companion paper (Ahlfeldt et al., 2021). In this setting we have full control over the data generating process, which allows us to observe the distribution of prime service firms in the historic period, surgically switch off agglomeration and transport channels, and abstract from threats to identification that originate from unobserved components in the error term or the endogeneity of transport networks.

4.1 Empirical evidence

The role of external returns in shaping multiple equilibria. As argued in Section 2.5, we expect external returns to be of subordinate relevance prior to 1900. Therefore, disasters before 1900 make for a useful placebo test. While we do not observe the distribution of prime services at the establishment level during the historic period, we can track the location of city halls, which serve as proxies for historic prime locations as documented in Figure 7. Table 5 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 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)). Due to the smaller sample, we abstain from using an instrument for historic population and put less weight on the parameter estimate, although the pattern is consistent with the absence of agglomeration-induced persistence in the first-nature geography era.

Table 5: The effects of city size and disasters on city hall relocations

	(1) Ln distance between first and 1900 city hall	(2) Distance between first city hall and 1900 city hall	(3) Ln distance between 1900 and 2000 city hall	(4) Distance between 1900 and 2000 city hall
Estimator	OLS	Poisson	OLS	Poisson
Ln population 1800	-0.039	0.226		
	(0.04)	(0.16)		
Disasters until 1900	-0.066	0.130		
	(0.06)	(0.14)		
Ln population 1900	0.032	-0.268	-0.132*	-0.219
	(0.08)	(0.23)	(0.06)	(0.17)
Disasters since 1900			0.133**	0.241***
			(0.04)	(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: Unit of observation is cities that existed in 1800. Geo. controls are the same as in Table 4. For details on the underlying data, see Global Cities dataset appendix (Ahlfeldt et al., 2020). * p < 0.1, ** p < 0.05, *** p < 0.01

Transport-induced persistence. City size-induced persistence in internal city structure can arise from durable transport networks—that are endogenous to historic city size due to substantial sunk and fixed costs—because these create a permanent accessibility advantage for the historic centre. Table 3 suggests that this mechanisms matters.

To disentangle the network channel from other channels that operate through city size, we allow for an independent transport network effect on effective concentration and expand expression (2) as follows:

$$C_i^t = c_i^t \times e^{a_i^t} \times (P_i^t)^{\theta^t} \times e^{\delta^t M_i^t}, \tag{5}$$

where M_i^t is an indicator for city i having a subway in period t and δ^t is the respective effect. Substituting (5) into (1) and regrouping terms, we obtain:

$$C_i^2 = \alpha \times (C_i^1)^{\zeta} \times (P_i^1)^{\beta-\zeta\theta^1} \times (P_i^2)^{\theta^2} \times e^{\gamma D_i^2 + a_i^2 - \zeta a_i^1 + \epsilon_i} \times e^{\delta^1 M_i^1 + \zeta\delta^2 M_i^2}. \tag{6}$$

Proceeding as in Section 3.1, our empirical specification is given by:

$$\ln C_i^2 = \widetilde{\alpha} + \widetilde{\beta} \ln P_i^1 + \delta^1 M_i^1 + \theta^2 \ln P_i^2 + \gamma D_i^2 + \widetilde{b} X_i + \widetilde{\delta}^2 M_i^2 + e_i, \tag{7}$$

with $\widetilde{\delta}^2 = \zeta\delta^2$. The mapping from the other reduced-form parameters and the error term to the structural parameters and fundamentals remains as in (4). We choose 1910 as the cutoff year for the subway in the historic period to allow for the time lag between planning and implementation of subway systems. This way, we capture that the decision to build a subway was likely made around 1900. Given the stylized evidence in Section 2, we expect a positive historic effect of early subway adoption, $\delta^1 > 0$.

Complementary to our discussion in Section 3.1, 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.5). 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

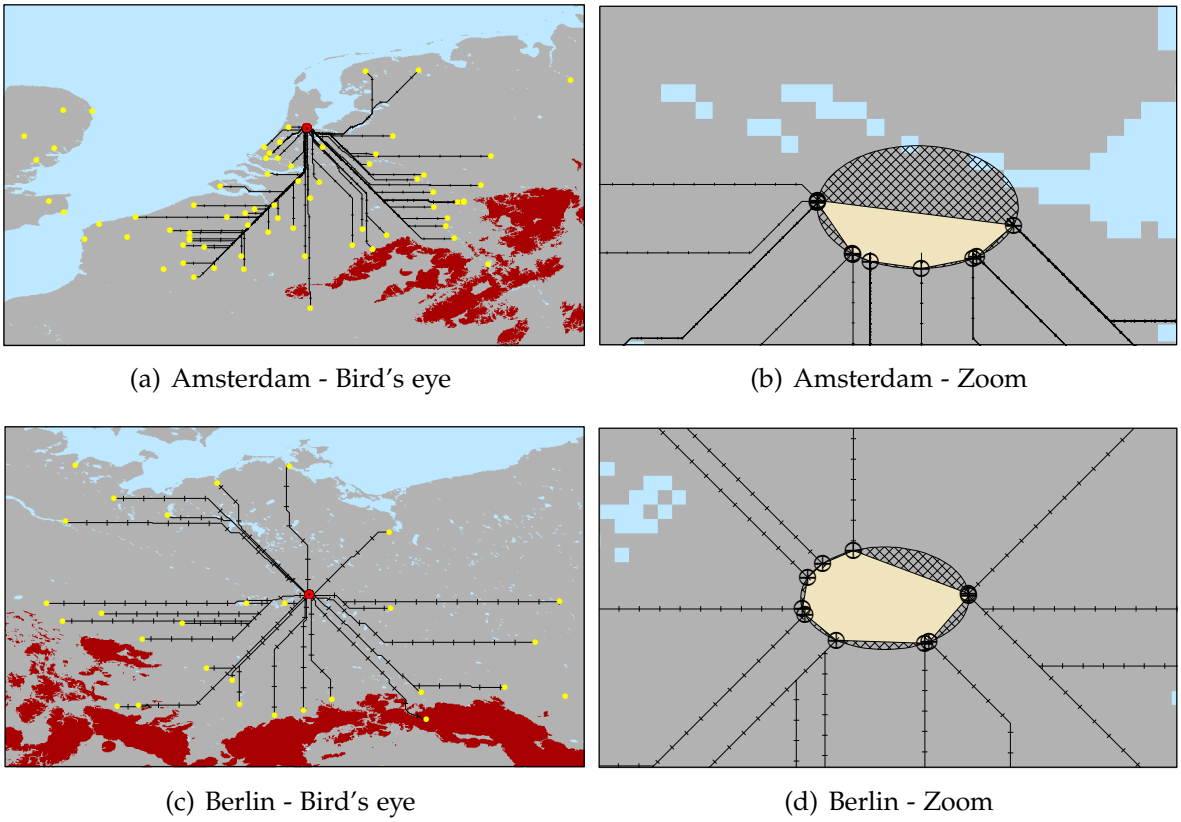
system after crossing the threshold? Besides the size of the city, the level of congestion also depended on a second factor. The number of unconnected ports of entry and exit, namely railway termini, the distance between them, and their spatial configuration were crucial for the city's congestion—and 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.[22] We keep the neighboring cities only if their distance remains below the equivalent of 300 kilometers by land transportation. Panels (a) and (c) of Figure 8 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 constrains (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.5). 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 8 provide corresponding illustrations for Amsterdam and Berlin. Despite having crossed the population threshold to build a rapid transit system at the dawn of the 20^{th} century, Amsterdam had a much lower probability to do so than Berlin according to our measure of subway potential. This is corroborated by the historical record. Whereas Berlin had adopted rapid

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

Figure 8: Subway potential of Amsterdam and Berlin



(a) Amsterdam - Bird's eye

(b) Amsterdam - Zoom

(c) Berlin - Bird's eye

(d) Berlin - Zoom

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.

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 fundamentals, 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).[23] This allows to disentangle the effect of durable transportation networks from other mechanisms captured by historic city size, such as agglomeration-induced persistence or a historic downtown building stock that could be favorable to contemporaneous prime services.

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

Our instrument now satisfies all exogeneity conditions, but for it to be relevant we have to model one more determinant. Consistent with the high fixed costs of a subway system, descriptive empirical evidence (see Appendix A.5) suggests 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.9.1 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 instrument, whereas such an interaction does not exist for the population instrument. We document this and the power of the instrument in Appendix A.9.1.

Table 6: Network-induced inertia

	(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 (×-1)	Ln dist. from EWPP to 1900 CH (×-1)	Ln mean dist. from all PLs to 1900 CH (×-1)	Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.097***	0.207***	0.215***	0.088***	0.144***	0.291***	0.320***
	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)	(0.07)	(0.09)
Subway in 1910	0.465***	0.465**	0.724***	0.526***	0.507***	0.572*	0.535+
	(0.14)	(0.23)	(0.27)	(0.14)	(0.15)	(0.29)	(0.34)
Disasters since 1900	-0.103***	-0.122**	-0.190***	-0.128***	-0.114***	-0.184***	-0.084
	(0.04)	(0.06)	(0.06)	(0.04)	(0.04)	(0.06)	(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., 2020) 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.9.1. Robust standard errors in parentheses. + $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6 presents estimates of specification (7) 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 with distance increases by 53 percentage points (column

4). The average distance between 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 4, and consistent with our priors, the estimated historic population elasticities capturing the residual channels 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.9.2). 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.7). On balance, our interpretation of the evidence is that size-induced persistence in the internal organization of cities originates at least partially from durable transport networks that were endogenous to city size a century ago. The remaining historic city size effect may capture aspects of durable transport networks not observed in data or other mechanisms such as agglomeration-induced persistence.

4.2 Monte Carlo study

Even when exploiting our data set to its fullest, the identification of mechanisms that moderate the causal relationships uncovered in Section 3 asks a lot of our data. There is no natural experiment that allows exploring the effect of natural disasters during the second-nature geography era in the absence of external returns. The separate identification of the subway and agglomeration channels is local by nature since few cities developed subway systems and those which did were all among the largest. Therefore, we now leave the empirical world to take control of the data generating process. Our aim is to replicate the empirical analysis in Table 6 in a world in which we can ensure that spatial concentration of prime services is solely shaped by the interplay of randomized variation in historic city size, transport networks, and spatial shocks as well as the canonical mechanisms in recent urban models.

The granular spatial model (GSM) we develop in a companion paper (Ahlfeldt et al., 2021) is particularly suitable for the task. The model follows the basic structure of the canonical quantitative spatial model (QSM) in which external returns generate agglomeration of jobs, commuting costs lead to co-location of jobs and workers, and inelastically supplied land generates a dispersion force (Ahlfeldt et al., 2015). The novel feature of the GSM is that it not only features explicitly defined locations, but also explicitly defined workers and firms. This provides the necessary granularity to reproduce one-for-one our real-world spatial concentration measures that are based on establishment data. The GSM lends itself to rich heterogeneity, allowing us to model competition on land markets among firm types

such as manufacturing firms, local prime services firms, and global prime services firms. Below, we briefly outline the GSM and discuss how we use it to run a series of Monte Carlo experiments to evaluate the long-run effects of temporary spatial shocks in the presence and absence of external returns as well as endogenously deployed mass transit systems. We conduct these Monte Carlo experiments within the actual geography of Chicago, which still represents the stereotype of the classic monocentric city (with most prime services concentrated in the "loop"). We show in Appendix A.10 that our parametrization of the model resembles the real-world Chicago in various outcomes (e.g., in terms of the distribution of land prices or commuting flows). There, we also provide details on the algorithm we use for the endogenous transport network formation (see our companion paper, Ahlfeldt et al. 2021, for details on the formal structure of the model and the simulation procedure).

The GSM in a nutshell. The basic building blocks of the GSM are locations, workers, and firms, which we distinguish into (local or global) prime services and manufacturing firms. As in the conventional QSM, there is a finite set of locations with exogenous land supply, productivity, and amenity value. Workers and firms behave similarly as in the standard QSM, but they exist as indivisible agents. Workers maximize a Cobb-Douglas utility from non-tradable floor space consumed at at spatially varying rent and a tradable good whose price is normalized to one. The wage they receive is deflated in effective commuting distance, which depends on the transport network. Firms maximize profits from producing either tradable prime services or manufactured goods whose prices are determined outside the city. A firm's productivity depends on the productivity of its location, which in turn depends on an exogenous component and an endogenous component that is a function of employment in nearby firms. Each worker requires a fixed amount of floor space to be productive. Floor space is produced by a competitive construction sector using inelastically supplied land and nationally traded capital. An equilibrium is characterized by land market clearing and the absence of any incentive to adjust location choices and the worker-firm matching. As with the canonical QSM, the primitives of the GSM consist of structural parameters and structural fundamentals (Redding and Rossi-Hansberg, 2017). Due to its granularity, the GSM allows for arbitrary heterogeneity in all model primitives at the level of locations, workers, and firms. We only parameterize the dimensions of heterogeneity that are most directly relevant to our analysis. For example, we use uniform canonical values for housing expenditure shares or commuting costs, but we require more floor space per worker for manufacturing firms, allow for greater returns to agglomeration for prime services firms, and let multi-establishment global firms produce spillovers that depend on a component that is external to the city (see Appendix A.10 for details).

Simulating the effects of disasters. To simulate the effects of random disasters, we conduct Monte Carlo experiments adhering to the following procedure. In the first step, we

simulate the *initial equilibrium* in a first-nature world where the external returns are set to zero. To this end, we take the distribution of location fundamentals as given, randomize the number of agents in the city (which directly corresponds to variation in city size), and randomly assign fundamental productivity to workers and firms. 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. In the second step, we enter the second-nature world and simultaneously: (i) turn on external returns; (ii) allow for the endogenous development of a transport network; and (iii) shock the center of economic activity 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, the larger cities tend to develop larger transport networks. We simulate a major disaster by enforcing the relocation of a random subset of 50-100% of the firms in the location with the largest prime service employment in the initial equilibrium. We find a *temporary equilibrium* where only the displaced firms find new optimal locations. In the third step, the shock dissipates and all firms re-optimise their locations and all workers re-optimise their assignment to firms. This delivers the *final equilibrium* where agglomeration economies, shocks, and transportation networks jointly shape the spatial structure of employment centres in the city.

These two sets of runs provide intensive- and extensive-margin variation in the shock and network variables that resembles our 'Global Cities' data set. Note that in all of these runs we turn on agglomeration economies in the final equilibrium. To gain further insights into the mechanisms, we, in addition, report results where we shut down agglomeration economies. More precisely, we simulate 1,000 runs without external returns but with shocks and network formation, and 1,000 runs with just shocks.

Analysis of simulation outputs. The estimation strategy outlined in Section 3.1 can be taken one-to-one to the artificial data set generated by the Monte Carlo experiments. Each Monte Carlo run produces aggregate outputs that we can index by i as they directly correspond to the different cities in our 'Global Cities' data set. The initial ($t = 1$) and the final ($t = 2$) equilibria directly correspond to the historic and the contemporary period in the data. Yet, we make subtle changes to (4) to leverage on our power to fully control the data generating process:

$$\ln C_i^2 = \widetilde{\alpha} + \zeta \ln C_i^1 + \widetilde{\beta} \ln P_i^1 + \gamma D_i^2 + \epsilon_i, \tag{8}$$

Since we do not change city size as we move from the first-nature to the second-nature world, we have $P_i^1 = P_i^2$ so that P_i^2 drops from the equation and the historic population effect becomes $\widetilde{\beta} = \beta - (\zeta\theta^1 - \theta^2)$. Since we observe effective concentration in the

initial equilibrium, C_i^1 remains in the estimation equation and $\zeta\xi_i^1$ drops from the error. Since we keep all observable and unobservable fundamental location characteristics constant across Monte Carlo runs, $(\rho^2 - \zeta\rho^1)f_i = (\rho^2 - \zeta\rho^1)\bar{f}$ and $\widetilde{b}X_i = \widetilde{b}\bar{X}$ enter the constant $\widetilde{\alpha} = \ln\bar{\alpha} + (\rho^2 - \zeta\rho^1)\bar{f} + \widetilde{b}\bar{X}$. The remaining error ϵ_i comes purely from the randomness in the initial allocation of productivity across agents, the initial allocation of agents to locations, and the randomness in the sequence in which agents make decisions. The beauty of this synthetic data set is twofold: First, we have exogenous and truly independent variation in disasters and historic population and a genuinely random error term. Second, we observe our artificial cities in four different worlds in which there exist either agglomeration economies or endogenous network formation, both, or neither.

Table 7: Origins of inertia

	(1) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
(a) Agglomeration economies ON, network ON							
Log city size	0.466***	0.187***	0.102***	0.578***	0.070***	0.388***	0.394***
	(0.017)	(0.013)	(0.027)	(0.026)	(0.018)	(0.040)	(0.040)
Shock intensity	-0.209***	-0.158***	-0.221***	-0.596***	-0.301***	-0.971***	-0.962***
	(0.023)	(0.016)	(0.035)	(0.037)	(0.024)	(0.058)	(0.058)
(b) Agglomeration economies OFF, network ON							
Log city size	0.586***	0.705***	0.083***	0.120	0.008	0.103**	0.087*
	(0.011)	(0.015)	(0.032)	(0.082)	(0.009)	(0.041)	(0.046)
Shock intensity	0.030**	-0.017	0.338***	0.399***	-0.175***	-0.067	-0.039
	(0.013)	(0.020)	(0.039)	(0.110)	(0.013)	(0.053)	(0.056)
(c) Agglomeration economies ON, network OFF							
Log city size	0.080***	0.004	0.036	0.033	0.093***	-0.049	-0.038
	(0.015)	(0.013)	(0.034)	(0.022)	(0.020)	(0.040)	(0.039)
Shock intensity	-0.177***	-0.197***	-0.260***	-0.724***	-0.289***	-1.264***	-1.265***
	(0.019)	(0.014)	(0.035)	(0.030)	(0.023)	(0.054)	(0.053)
(d) Agglomeration economies OFF, network OFF							
Log city size	-0.002	0.003	-0.107***	-0.112	-0.008	-0.275***	-0.260***
	(0.007)	(0.010)	(0.034)	(0.078)	(0.008)	(0.040)	(0.043)
Shock intensity	0.031***	-0.024**	0.422***	0.377***	-0.171***	-0.073	-0.050
	(0.009)	(0.012)	(0.043)	(0.100)	(0.012)	(0.056)	(0.058)

Notes: Standard errors in parentheses. CH = city hall, EWWP = employment-weighted workplaces, PL = prime location (identified using the same approach as described in 2.4). Unit of observation is the output of one run of the GSM simulation model. 1,000 observations in all models. All models control for initial concentration. 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 taken as the location of the largest prime location in the initial equilibrium. Standard errors (in parentheses) are bootstrapped in 1000 iterations. $^{+}$ $p < 0.15$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7 summarises OLS regression results from (8) for these four different worlds. Panel (a) corresponds to the real world as we allow for agglomeration economies and endogenous network formation in the second-nature era. The results closely resemble our estimates based in our 'Global Cities' data set presented in Table 4. Our virtual disasters lead to a more dispersed spatial distribution of prime services and to contemporary prime locations further away from the historic prime location. Confirming the role of agglomeration economies as

facilitators of multiple equilibria in spatial configurations, shocks no longer disperse prime services when we disable external returns to scale in panel (b).Shocks have inconsistent effects on the spatial concentration of economic activity, depending on the measure used. Since the GSM features some frictions to the movement of agents, it is natural for shocks to have some effects on the equilibrium location of agents, even in the absence of agglomeration economies. In contrast, the city size effect persists, suggesting that agglomeration-induced persistence is not the main driver of inertia in our model. This interpretation is substantiated by panel (c) where, enabling agglomeration economies and disabling network formation, we no longer observe a significant city-size effect on any measure that captures prime services concentration at a *local* level. Hence, by creating a permanent advantage in the form of lower commuting costs around central nodes, durable networks in large cities play a critical role in anchoring prime locations. Finally, once we disable agglomeration economies and network formation in panel (d), the empirical relationship between initial city size and contemporaneous concentration of prime services we uncovered in our data vanishes altogether, lending further support to our hypothesised mechanisms.

The important lesson from the Monte Carlo experiments within the GSM is that a model with canonical assumptions regarding worker and firm behavior can generate the patterns we observe in our 'Global Cities' data set. Our findings substantiate the notion that external returns give rise to multiple equilibria in the spatial distribution of knowledge-based tradable services. Without dismissing alternative sources of inertia outside our model, we can rationalize the empirically observed fact that prime locations are more resilient in historically larger cities through the permanent location advantage endogenous durable transport networks deliver to historic prime locations.

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.

However, little is known about how these forces play out at a small spatial scale within cities. Our findings show that the structure of large cities is prone to multiple equilibria and inertia, 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 unless the intended new prime locations become themselves focal nodes within large urban transport networks.

References

Acemoglu, Daron, 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: Global Cities Data Appendix," 2020.

_ , _ , and _ , "A granular spatial model," 2021.

Allen, Treb and Dave Donaldson, "Persistence and Path Dependence in the Spatial Economy," Working Paper 28059, National Bureau of Economic Research November 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.

Bairoch, Paul, *Cities and economic development: From the dawn of history to the present*, University of Chicago Press, 1988.

Baum-Snow, Nathaniel, 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.

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.

Brueckner, Jan K., "Chapter 20 The structure of urban equilibria: A unified treatment of the muth-mills model," *Handbook of Regional and Urban Economics*, 1 1987, *2*, 821–845.

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.

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** __ , "The Economics of Urban Density," *Journal of Economic Perspectives*, 2020, *34* (3), 3–26.

__ **and Matthew A. Turner**, "Urban growth and transportation," *The Review of Economic Studies*, 2012, *79* (4), 1407–1440.

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.

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.

Fajgelbaum, Pablo D and Edouard Schaal, "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, jul 2020, *88* (4), 1411–1452.

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.

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.

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.

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.

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
FIRE: Finance, insurance, real estate, rental, and leasing				
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
PROF: Professional and business services				
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
$\sum$		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

	MSAs		Non-MSAs		Total
Prime service	M$	Share	M$	Share	M$
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:

$$\frac{\text{VA}_{sm}}{\text{VA}_m} = c_s + a_s \ln P_m + e_{sm},$$

where VA is value added, $s \in$ *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.[24] Figure 1 shows the results for prime services.

[24]For 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., 2020). 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.[25]

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 $\mathcal{P}^T_{l,c}$ 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 $\mathcal{P}^T_{l,c}$, we estimate the following empirical specification in city-specific regressions:

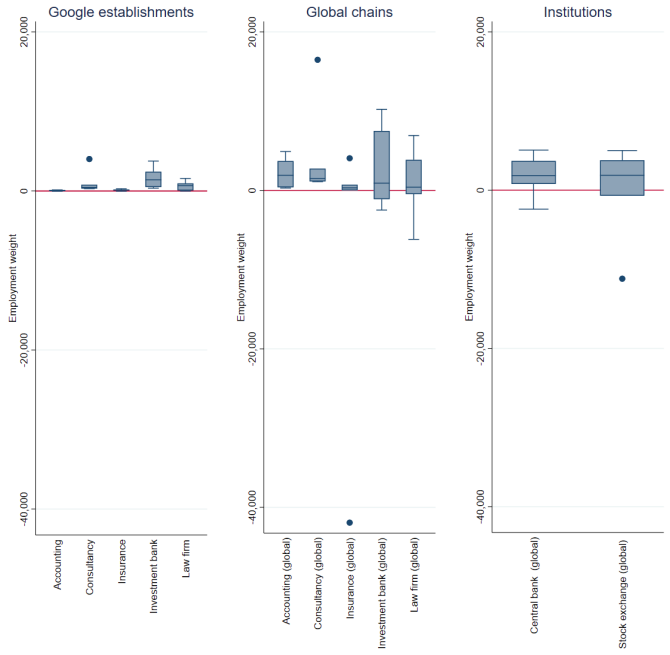
$$\mathcal{E}_{l,c} = a + \mathcal{P}^T_{c,l} b^T_c + \varepsilon^T_{l,c}, \tag{9}$$

where b^T_c are city-establishment-type-specific employment weights and $\varepsilon^T_{l,c}$ 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.[26] 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.

[25] We 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., 2020).

[26] There 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		

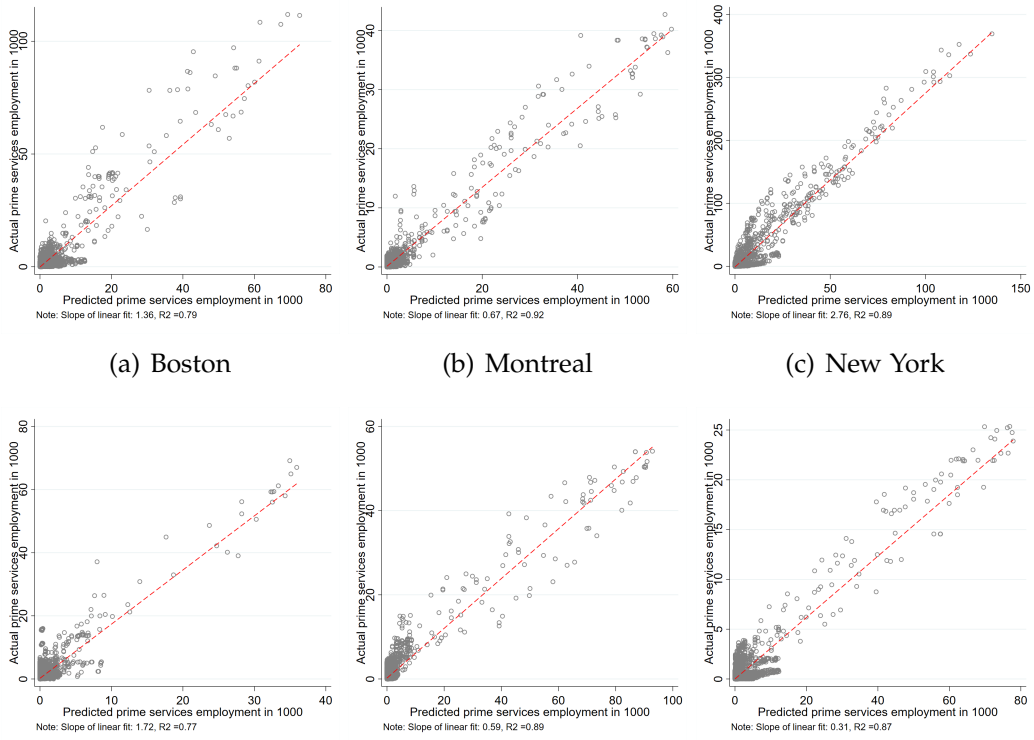
Notes: Employment weights from city-specific regressions of PS employment on PP counts (within random 750-m-radius disks). The values reported in the table are the median values in the distributions of estimated employment weights across cities by establishment type. Values are rounded to the nearest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

Since we wish to use the estimated employment weights $\widehat{b}^T$ 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 $\widehat{b}^T$ to generate a measure of predicted prime service employment $\widehat{\mathcal{E}}_{l,c} = \mathcal{P}_{c,l}^T \widehat{b}^T$. 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



(a) Boston (b) Montreal (c) New York

(d) Philadelphia (e) Toronto (f) Vancouver

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.

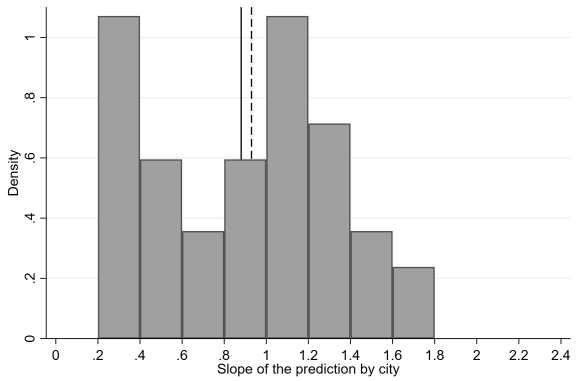
Table A 4: Overidentification of EWPPs

	(1) Prime places count	(2) SNL-S&P count	(3) Starbucks count	(4) Co-working count	(5) Actual PS emp. (1000)	(6) Actual PS emp. (1000)
Predicted PS employment (1000)	0.928***	0.371***	0.351***	0.205***	1.007***	0.493***
	(0.016)	(0.005)	(0.014)	(0.004)	(0.165)	(0.040)
Constant	0.091***	0.121***	0.010	-0.033***	0.067	0.997***
	(0.008)	(0.003)	(0.006)	(0.002)	(0.281)	(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.[27] 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.

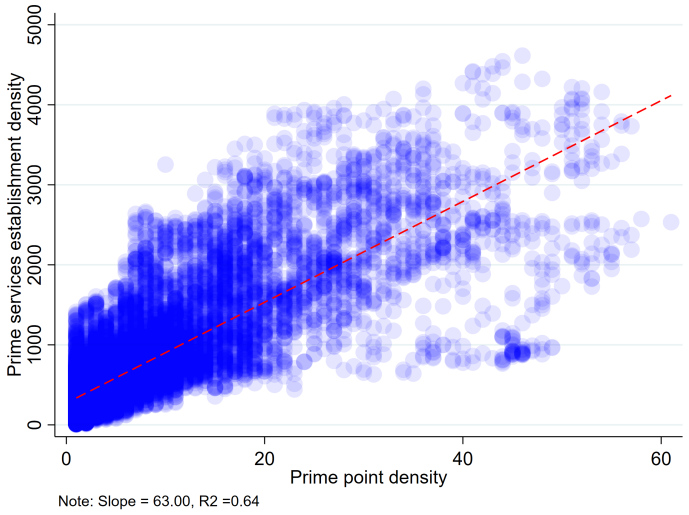
For overidentification tests outside the US, we distinguish between an extensive margin (can we predict the location of prime services establishments) and an intensive margin (can we predict the employment of prime services establishments). In Table 1, we have already shown at a global scale that prime points are spatially correlated with places where prime services employees work and meet (Grade-A offices, co-working spaces, Starbucks franchises), which points to predictive power at the extensive margin. Establishment-level prime services data is notoriously difficult to get hold of. But we do have access to a very large sample of establishments for Moscow. For a detailed extensive-margin overidentification test, we first generate 30,000 randomly drawn disk with a radius of 750 meters. We then count the prime points queried from Goolge Places and the actual prime services establishments within those disks. Finally, we correlate both count measures in Figure A4. There a positive correlation as expected. A simple regression of the prime point density on the prime services establishment density delivers an R^2 of 0.64. This result substantiates the impression that our prime points have strong predictive power at the extensive margin, also outside the US.

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.

[27] Dropping 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%.

Figure A4: Prime establishments vs. prime point in Moscow



Notes: This figure compares the observed prime service establishments from the SPARK/Inferfax database (covering 2012-2014) with the prime points queried using the Google Places API at the level of 30,000 750-radius-disks drawn around 30,000 random points..

Table A 5: Prime location specialization

Within 50 km of	NYC Empire State	Boston Prudential	Philly Liberty Bell	Toronto CN Tower	Montreal KPMG Tower	Vancouver MNP Tower	Mean 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., 2020). 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.

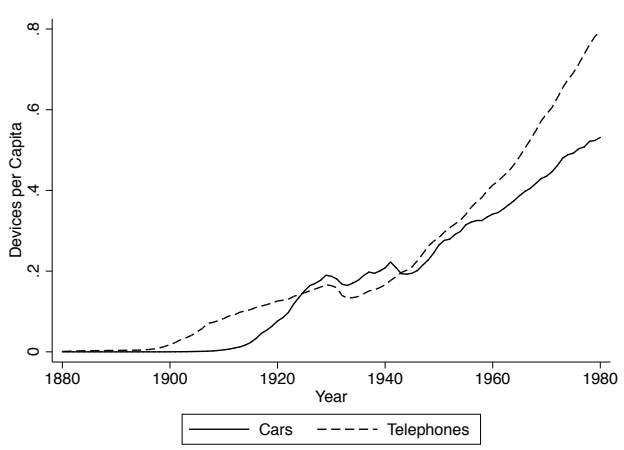
A.4 Measures of spatial concentration

Table A 6: Measures of spatial concentration

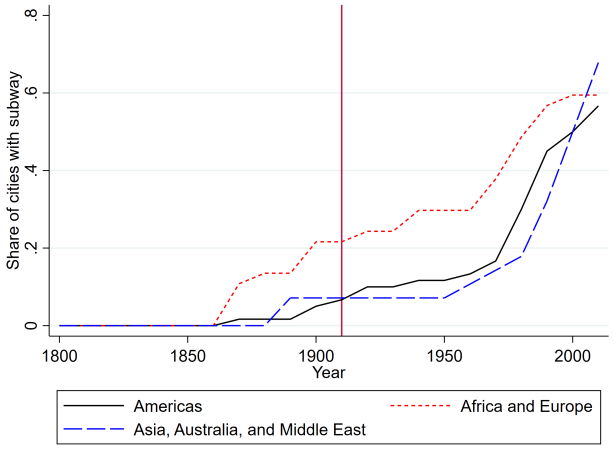
Measure	Shade of primeness	Formula	Notation & comments
		Modern concentration measures	
Bilateral distances between...			
...EWWP (weighted)	General prime service sprawl	$G = \frac{1}{\sum_i \sum_{k \neq i} E_i E_k} \sum_i \sum_{j \neq i} d_{ij} E_i E_j$	d_{ij}: bilateral distance between two prime points 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.
		Concentration relative to early center	
Central point	Direction (N/E/S/W) of center relocation (visual)	$\arg\min_i G_i = \frac{1}{\sum_{k \neq i} E_i E_k} \sum_{j \neq i} d_{ij} E_i E_j$	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	$\overline{d_m} = \frac{1}{N_m} \sum_i^{N_m} d_{im}$	m: city hall 1900; i: EWWP location
...all PLs	Attraction of old center for new centers	$\overline{d_m} = \frac{1}{N_m} \sum_i^{N_m} d_{im}$	m: city hall 1900; i: prime locations
...global PLs	Persistence of most important centers	$\overline{d_m} = \frac{1}{N_m} \sum_i^{N_m} d_{im}$	m: city hall 1900; i: global prime locations

A.5 First nature, second nature, and transport revolution

Figure A5: Transport and telecommunication through the ages

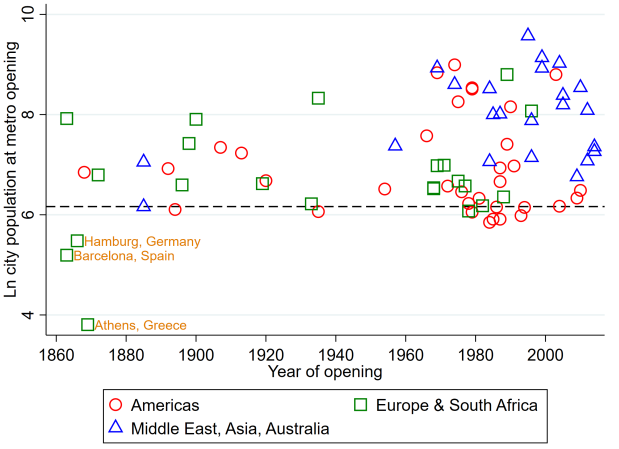


(a) Cars and Telephones per capita (United States)



(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. (2020) for underlying data sources.

The Transport and Communication Revolution. Panel 4(a) of Figure A5 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 A5 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 A5 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:

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

Based on data on population densities *before* the advent of rapid public transport, we can derive a plausible maximum population density $\text{POPDENSITY}_{\text{max}}^{\text{city}}$ from historical data and then solve for the population threshold $\text{POP}_{\text{threshold}}$.[28]

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

[28] Bairoch (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).

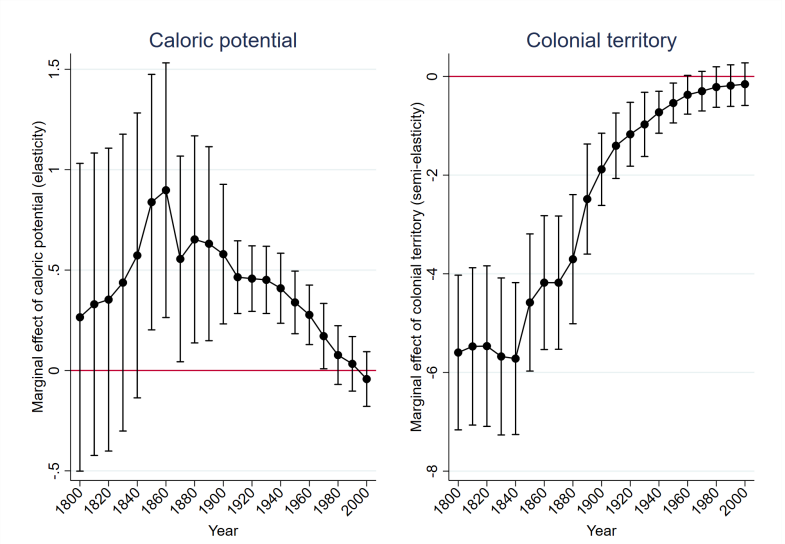
increase as inhabitants would need to overcome distances by means other than walking. They would then request public transport.

Does this threshold correspond to the observed introductions of rapid urban transport systems? Panel 4(d) of Figure A5 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.

A.6 Instruments for population

Figure A6 shows that our instrumental variables *caloric potential* and *colonial territory* are significantly correlated with city population in 1900, but not in 2000.

Figure A6: 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., 2020).

In Table A 8, we correlate the 1900 height gradient by city with disasters, population, and the instruments for the latter. The 1900 height gradient is the slope coefficient of a city-specific regression of the log of building height in 1900 against the log of distance from the city hall in 1900. In short, we find no significant correlation.

Table A 8: 1900 city hall gradients and population instruments

	(1) 1900 height gradient	(2) 1900 height gradient	(3) 1900 height gradient	(4) 1900 height gradient	(5) 1900 height gradient	(6) 1900 height gradient	(7) 1900 height gradient
Ln population 1900	0.095				0.094		
	(0.11)				(0.10)		
Disasters since 1900		0.039			0.021	0.054	0.057
		(0.10)			(0.08)	(0.11)	(0.11)
Ln average caloric potential			-0.291			-0.277	-0.303
			(0.21)			(0.22)	(0.23)
Occupied territory in 1800				0.164		0.119	
				(0.10)		(0.07)	
Observations	67	67	67	67	67	67	67
R^2	.0521	.00258	.0424	.0182	.0528	.0572	.048

Notes: Unit of observation is cities. Dependent variable is the slope parameter from a city-specific regression of log 1900 building height against log distance from the 1900 city hall. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A.7 Empirical results - OLS estimates

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

	(1) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.139***	0.216***	0.249***	0.103***	0.167***	0.265***	0.247***
	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)	(0.05)	(0.07)
Disasters since 1900	-0.117***	-0.118**	-0.201***	-0.133***	-0.122***	-0.168***	-0.049
	(0.04)	(0.06)	(0.06)	(0.03)	(0.04)	(0.06)	(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 4.

Table A 10: Network-induced inertia - OLS

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

A.8 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) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.120*	0.282***	0.269***	0.216***	0.181**	0.386***	0.389**
	(0.06)	(0.09)	(0.10)	(0.06)	(0.08)	(0.14)	(0.16)
Disasters since 1900	-0.116***	-0.123**	-0.202***	-0.141***	-0.123***	-0.176***	-0.059
	(0.03)	(0.05)	(0.06)	(0.04)	(0.04)	(0.06)	(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 4.

Table A 12: Robustness of disaster effects

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

Notes: See Table 4.

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 4). To keep the coefficients' magnitudes comparable, we instrument all disasters with the count of disasters conditional on the respective restriction.

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.

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

Table A 13: Sources of variation: within-country

	(1) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.082***	0.243***	0.131***	0.252***	0.098***	0.238***	0.237***
	(0.02)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Disasters since 1900	-0.056**	-0.047	-0.154***	-0.057	-0.055**	-0.069	0.008
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)	(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 4.

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) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.151***	0.202***	0.319***	0.103***	0.182***	0.303***	0.259***
	(0.04)	(0.05)	(0.08)	(0.03)	(0.04)	(0.08)	(0.09)
Disasters since 1900	-0.192	-0.215	-0.329*	-0.353***	-0.221+	-0.380**	-0.235+
	(0.15)	(0.19)	(0.18)	(0.11)	(0.15)	(0.18)	(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 4.

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 $I(\text{NA} = 1)$ or the rest of the world $I(\text{NA} = 0)$, 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) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
NA=0 × Ln population 1900	0.211***	0.269**	0.474***	0.100	0.302***	0.407**	0.327*
	(0.08)	(0.12)	(0.14)	(0.08)	(0.09)	(0.17)	(0.17)
NA=1 × Ln population 1900	0.085**	0.205***	0.145**	0.207***	0.101**	0.197**	0.220*
	(0.04)	(0.08)	(0.06)	(0.08)	(0.05)	(0.10)	(0.12)
NA=0 × Disasters since 1900	-0.079**	-0.091+	-0.181**	-0.171***	-0.089**	-0.175***	-0.077
	(0.04)	(0.06)	(0.07)	(0.03)	(0.04)	(0.06)	(0.07)
NA=1 × Disasters since 1900	-0.104***	-0.086	-0.186**	-0.037	-0.113**	-0.091	0.011
	(0.04)	(0.08)	(0.08)	(0.07)	(0.05)	(0.11)	(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 4 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., 2020).

A.9 Transport induced persistence

A.9.1 Baseline 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: $\mathcal{Y}_i = c_l^{\mathcal{Y}} + b_l^{\mathcal{Y}} I(CT_i = 1) + \varepsilon_{i,l}^{\mathcal{Y}}$, where $\mathcal{Y} \in \{\ln P^1{}_i, M_i^1\}$, $I(.)$ 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)).

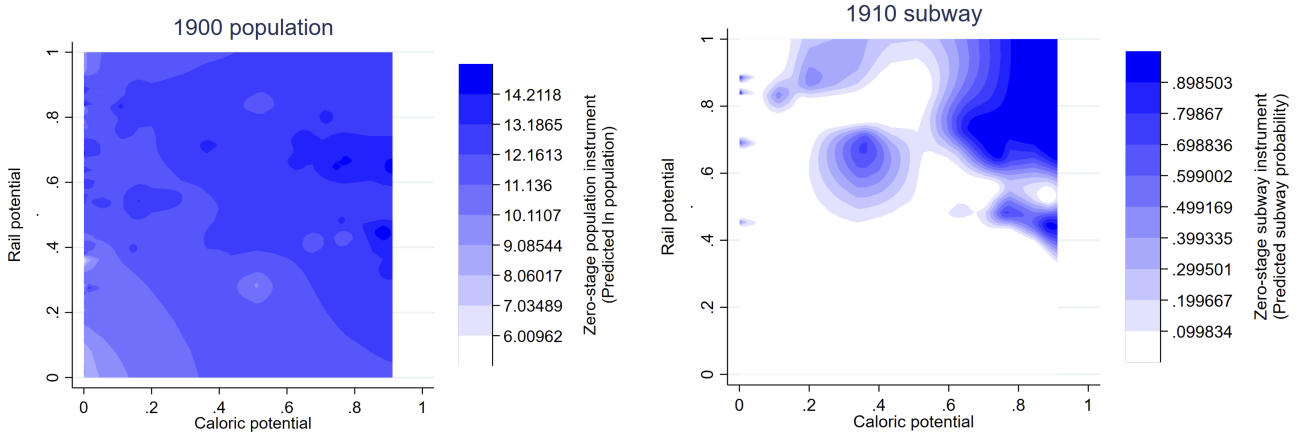
Table A 16: Relevance of zero-stage IVs

	(1) Log city population 1900	(2) Log city population 1900	(3) Log city population 1900	(4) Subway in 1910	(5) Subway in 1910	(6) Subway in 1910
Zero-stage population IV	1.060***		0.965***	-0.010**		
	(0.18)		(0.19)	(0.00)		
Zero-stage subway IV	0.067			1.227***		1.234***
	(0.53)			(0.07)		(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

The population instruments should predict historic population monotonically and historic subways only through an interaction with subway potential. Figure A7 illustrates how the exogenous historic population and subway shifters enter the zero-stage instruments.

Figure A7: 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.9.2 Alternative zero stages

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 $I(CT_i)$: $\mathcal{Y}_i = \sum_{u=0}^{B}\sum_{v=0}^{B}\sum_{w=0}^{1} b_i^{\mathcal{Y},u,v,w}\Big((CP_i)^u \times (SP_i)^v \times I(CT_i = w) + \varepsilon_i^{\mathcal{Y}}\Big)$, 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) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.085*	0.235***	0.269***	0.112***	0.147***	0.390***	0.431***
	(0.05)	(0.07)	(0.07)	(0.04)	(0.05)	(0.08)	(0.10)
Subway in 1910	0.758***	0.489	1.090**	0.649***	0.717***	0.535	-0.027
	(0.20)	(0.34)	(0.45)	(0.17)	(0.22)	(0.43)	(0.52)
Disasters since 1900	-0.098***	-0.127**	-0.196***	-0.131***	-0.113***	-0.200***	-0.106
	(0.04)	(0.06)	(0.07)	(0.04)	(0.04)	(0.07)	(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 $B = 2$.

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

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

Correspondingly, the two instrumental variables are:

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

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

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) Ln distance between EWPPs x (-1)	(2) Ln CD of bilateral EWPP distances at 0.75 km	(3) Ln average bilateral distance between PLs (x-1)	(4) Distance from 1900 CH gradient (×-1)	(5) Ln dist. from EWPP to 1900 CH (×-1)	(6) Ln mean dist. from all PLs to 1900 CH (×-1)	(7) Ln mean dist. from global PLs to 1900 CH (×-1)
Ln population 1900	0.150**	0.303***	0.261***	0.155***	0.226***	0.462***	0.519***
	(0.06)	(0.09)	(0.09)	(0.05)	(0.07)	(0.10)	(0.13)
Subway in 1910	0.932***	0.740	1.658**	0.633**	0.819**	0.495	-0.158
	(0.34)	(0.54)	(0.69)	(0.30)	(0.35)	(0.59)	(0.69)
Disasters since 1900	-0.107***	-0.135**	-0.190***	-0.138***	-0.125***	-0.212***	-0.122+
	(0.04)	(0.06)	(0.07)	(0.04)	(0.04)	(0.07)	(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.

A.10 Monte Carlo approach

We use a Monte Carlo approach to generate a synthetic version of the 'Global Cities' data set that we subject to same econometric specifications as the real-world data. In doing so, we simulate a version of the granular spatial model developed in a companion paper (Ahlfeldt et al., 2021). In this appendix, we show how we augment the model to account of endogenous transport network formation and how we parametrize the model.

A.10.1 Parametrization

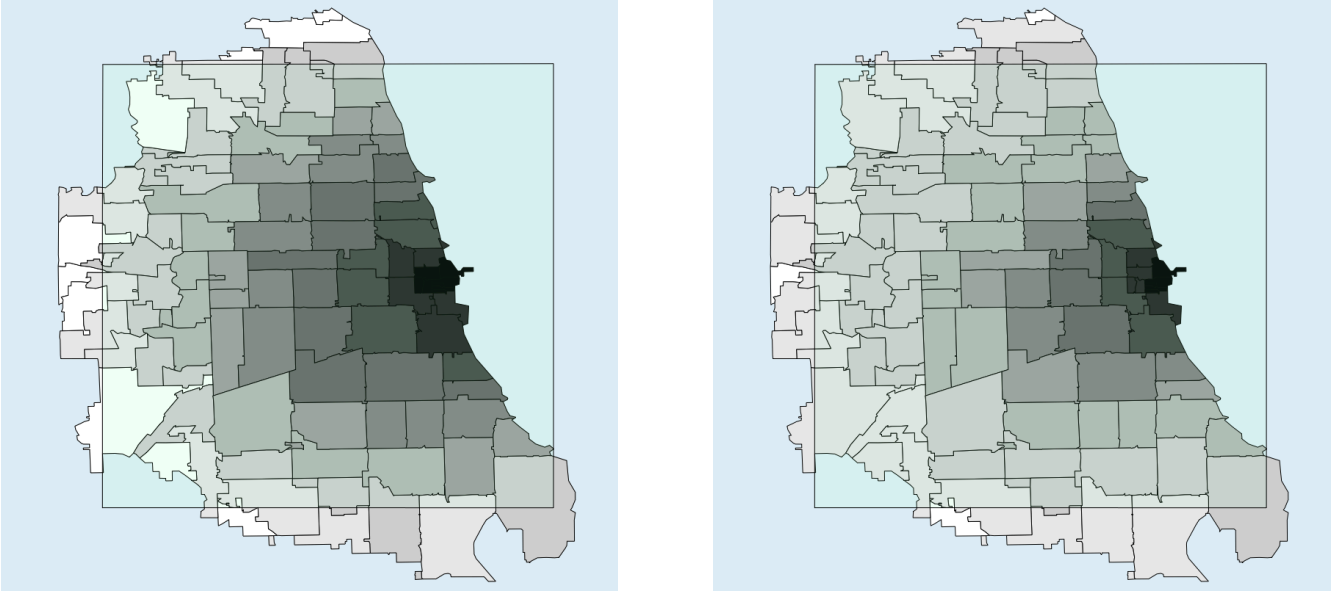
Except of the added endogenous transport network, we work with GSM that is identical to the one developed in Ahlfeldt et al. (2021). To save space, we do not replicate the development of the model here. Instead, we focus on how we parametrize the model using exactly the same notations as in Ahlfeldt et al. (2021). The main purpose of this section is to illustrate in an intuitive way that we conduct our Monte Carlo experiments within a simulated city that resemble the spatial structure of a real-world city.

A.10.1.1 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 A11 shows the extent of our empirical grid).

Figure A8: Distribution of production amenities and population

(a) Production amenities (b) 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: $B_\ell = \mathrm{e}^{-0.01 \times d_{\ell,FP}} \times B_{FP}$. 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:

$$\text{pop}_\ell = \frac{\mathrm{e}^{-0.1 \times d_{\ell,FP}}}{\sum_\zeta \mathrm{e}^{-0.1 \times d_{\zeta,FP}}} \times \text{total pop} \quad \text{and} \quad \text{firms}_\ell = \frac{\mathrm{e}^{-0.2 \times d_{\ell,FP}}}{\sum_\zeta \mathrm{e}^{-0.2 \times d_{\zeta,FP}}} \times \text{total firms}. \tag{10}$$

Since agents are indivisible in the GSM, 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.[29] Panel (a) of Figure A8 shows the spatial structure of our production amenities B_ℓ, while panel (b) displays the allocation of workers. This initial allocation is fairly monocentric.

[29]The total population is set at total pop $= 824$ in our illustrative example. Given rounding, we have 268 (potential) firms; these are broken down into 19 global prime service firms, 169 non-global PS firms, and 80 manufacturing firms. In the Monte Carlo simulations, we randomly draw a population size between 500 and 1000 from a uniform distribution to generate variation in city size. Prime service firm are 70% of the firms, with the remaining 30% being MFG firms. Within the PS firms, 10% are global.

A.10.2 Heterogeneous agents

Workers and firms are heterogeneous. They each get a productivity draw θ_ω (workers) or θ_φ (firms) from a uniform distribution. Firms draw from $\theta_\varphi \in [5, 8.5]$, whereas workers draw from $\theta_\omega \in [0, 5]$. 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 5 for non-global firms, whereas global firms are modeled as having a better span-of-control that ranges uniformly from 2 to 7. The SOC elasticity is set to 1.4 for all firms.[30] Each (local or global) prime service firm requires 2.1 units of office space per worker, whereas a manufacturing firm requires 33% more space per worker. We impose no correlation between θ_φ and $\overline{R}_\varphi$, i.e., firms differ along two independent dimensions in their productivity.

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 791 assigned workers. Hence, 33 workers remain unemployed (get zero wage) and 82 potential firms (seeds) cannot attract workers (end up being inactive). Obviously, by the matching process it is the low-productivity workers who remain unemployed (average productivity of 0.103, compared to 2.62 for the employed workers); and the low-productivity firms that do not operate (average productivity of 5.640, compared to 7.149 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 A9). Assortative matching leads (on average) to a convex relationship between productivity and wages (see panel (b) of Figure A9).

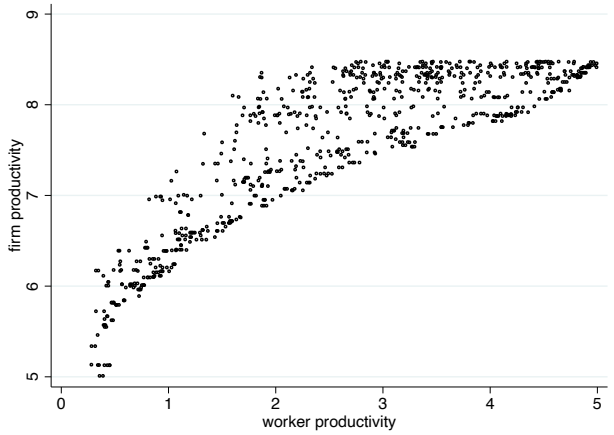
Turning to the firm-size distribution, it is depicted by panel (c) of Figure A9. The matching process and our parameterization yield a distribution of firm sizes with many small firms and a few large players, a realistic feature of observed firm-size distributions that are close to Pareto. A simple regression of log rank on log size yields a slope of -1.13 (standard error 0.037 and adjusted R^2 of 0.834) in the initial equilibrium assignment. The skewness in firm sizes maps into skewed profits (see panel (d) of Figure A9). The aggregate value of output produced in the example is 18,339.55, of which 6052 (one third) go to firms' profits and the remaining two-thirds are paid to workers. Workers' income net of commuting costs is and 11,247.79 (61.3% of the two-thirds wage bill), i.e., about 91% of their gross income. Thus, aggregate commuting costs are about 9% of gross income, a reasonable number.[31]

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

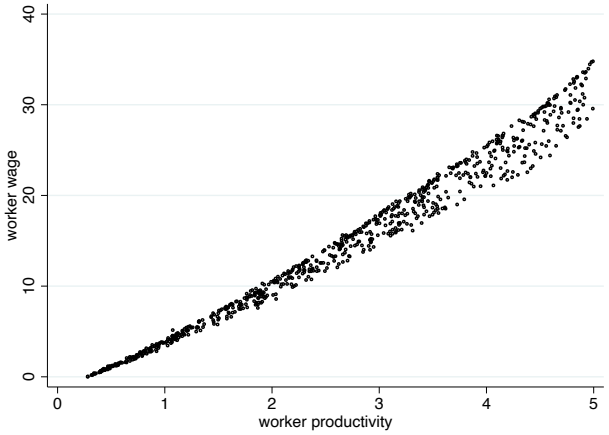
[31] According to BLS data for 2019, US consumers spent on average $63,036 ('Average annual expenditures') of which $10,742 was dedicated to commuting and other trips ('Transportation') (see `https://www.bls.gov/news.release/cesan.nr0.htm` for the data). This represents about 17% of income. The 2017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

Figure A9: Fuzzy assortative matching, productivity, firm size, profits, and wages

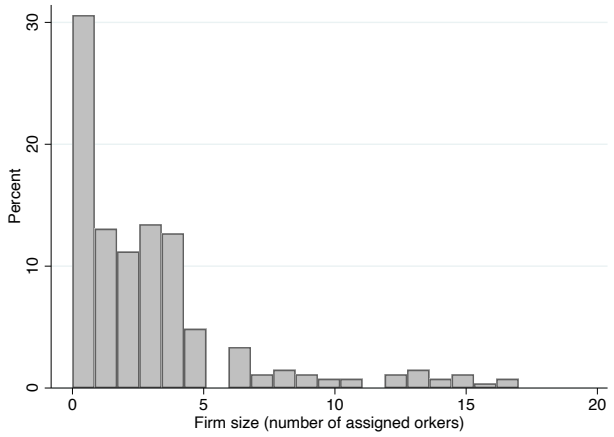
(a) Fuzzy assortative matching



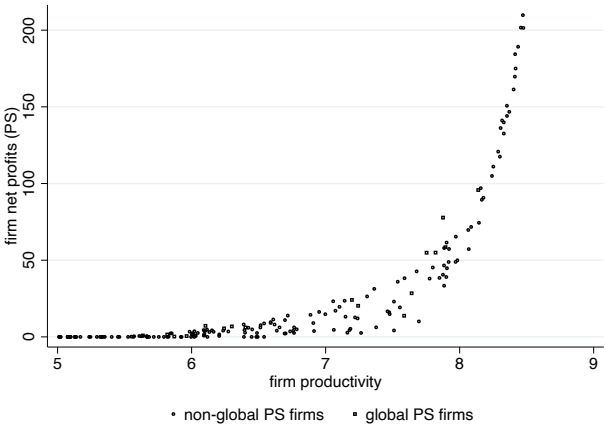
(b) Productivity and wages



(c) Distribution of firm sizes



(d) Productivity and profits, PS firms



Notes: Sample run of the ABM. Panel (a) depicts the relationship between firm and worker productivity and shows 'fuzzy' assortative matching. Panel (b) depicts the relationship between worker productivity and wages. Panel (c) depicts the histogram of firm sizes as measured by the number of assigned workers. Last, panel (d) illustrates the relationship between firm productivity and net profits, for non-global firms (circles) and global firms (squares). Since global firms have better SOC, they have higher profits conditional on productivity.

A.10.3 Transport network formation and commuting

Cities may endogenously develop transportation networks that will be used by workers for commuting. A network requires to pay a fixed setup cost (10,000); a cost per station (node) of the network (200); and a per kilometre cost of building links between nodes (50). The network is financed by a uniform wage and profits tax that is levied on all agents in the initial equilibrium (15%). The network is then constructed using the following procedure.

- From the worker-firm assignment, we compute all bilateral commuting flows between locations in the city. The link with the largest flow seeds the network, i.e., locations ℓ

and j with the largest flow get the first link and the first two nodes. Stations (nodes) are the access points to the network. We assume they are constructed at the center of the corresponding ZCTA.

- We take at the second largest flow for developing the second link. We do not allow for a new station to be built at less than some threshold distance (here, 1 kilometre) from an already existing station.
- We check whether the new link is cheaper to build by attaching it to the closest existing node of the network rather than by building a new direct link between the origin and destination.
- The procedure is repeated as long as budget remains. If the cost of a link exceeds the remaining budget so that the link cannot be built, we move to the next largest positive commuting flow until either no more budget is available or all links with positive flows have been checked (and potentially build).

Once the network is formed (recall this need not be the case), agents optimize by choosing whether to commute directly or use the network. In other words, there is modal choice. Without a network, agents move along the straight-line distance to firms. Let $d_{\ell j}$ denote the distance between locations ℓ and j. In the presence of a network, the distance is

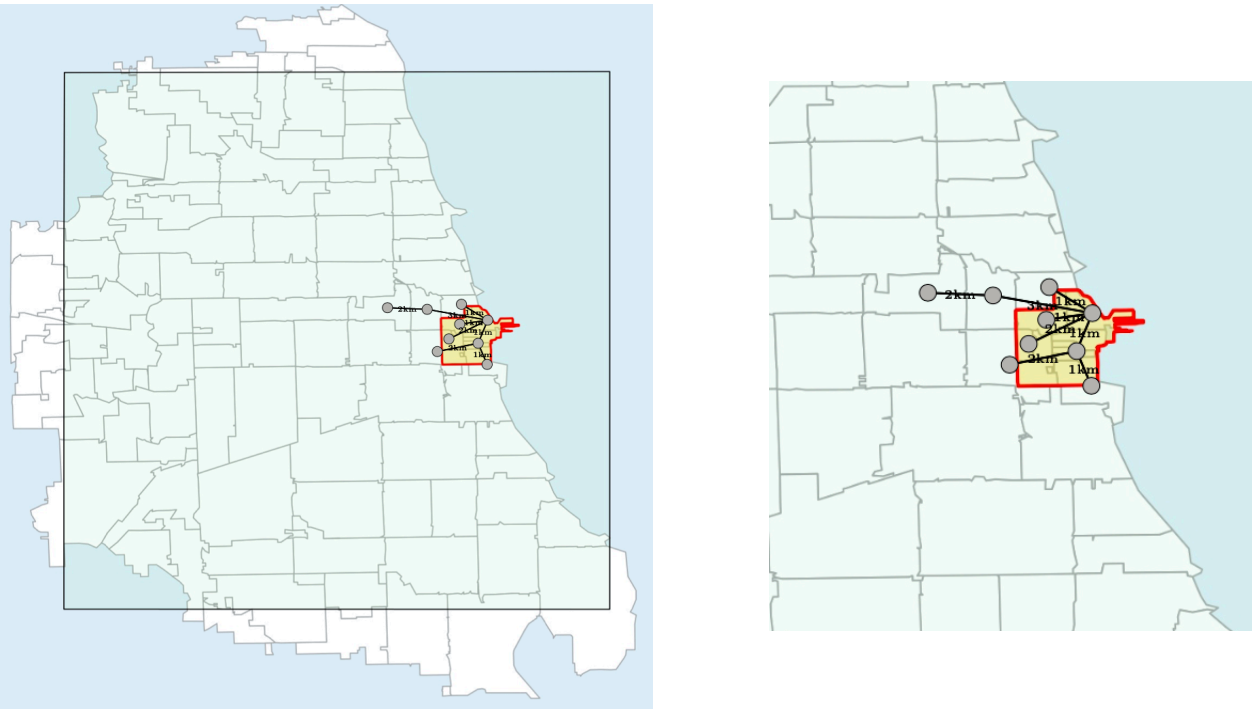
$$d^N_{\ell j} = \min_{n_1 \in \mathcal{N}} d_{\ell, n_1} + \min_{n_2 \in \mathcal{N}} d_{n_2, j} + \zeta \times \text{shortest path}(n_1, n_2), \tag{11}$$

where shortest path(n_1, n_2) is computed using a recursive implementation of Dijkstra's shortest path algorithm. The parameter $0 < \zeta < 1$ captures the travel time gains once on the network (if $\zeta \geq 1$, moving on the network is by construction never profitable given the triangle inequality). As seen from (11), access to and from the network is straight-line, and then there is a gain from moving along the network (distance on the network is equivalent to only ζ times distance off the network). Of course, agents choose the network if and only if $d^N_{\ell j} < d_{\ell j}$, and they travel off the network otherwise. For simplicity, we disregard aspects such as network pricing and user fees.

Figure A10 illustrates the network that is formed in our example run used in this appendix. As shown, a network with 9 stations and total length of 16.22 kilometers is developed. Panel (a) of Figure A10 shows that the network is, by construction, centered on the CBD that has the largest commuting flows. Panel (b) zooms in to show the structure of the network in more detail.

Observe that because employment is highly concentrated in the city, the commuting flow matrix is dominated by zeros, a realistic feature. In our initial equilibrium, 96.61% of the 10,302 ($= 102 \times 101$) bilateral commuting flows are zero. Among the non-zero flows, the top-3 destinations account for more than 70% of the commutes. This provides a strong

Figure A10: Example of endogenous transport network formation



Notes: Sample illustration of network formation, final equilibrium.

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.[32]

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).

A.10.4 Equilibrium

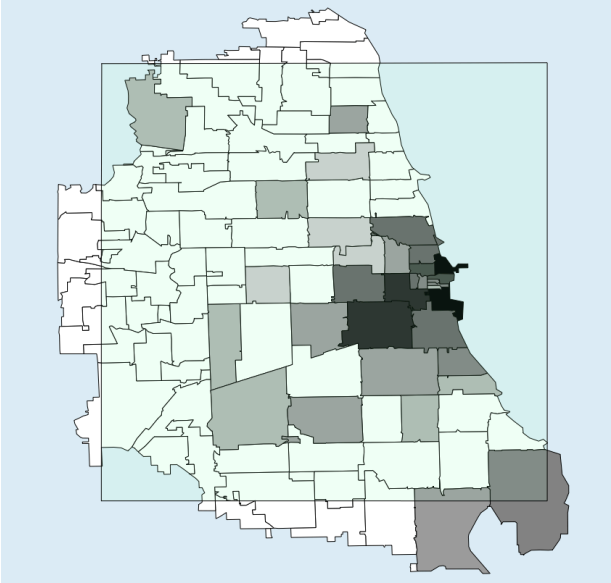
An equilibrium is such that: (i) workers optimally choose firms; (ii) firms optimally choose locations;[33] and (iii) rents clear the real estate market in all locations. An example of an initial equilibrium distribution of employment is shown in panel (a) of Figure A11. It is largely concentrated in the historic center of the city (where the productivity advantage is

[32]We 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 0.3 kilometres 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.

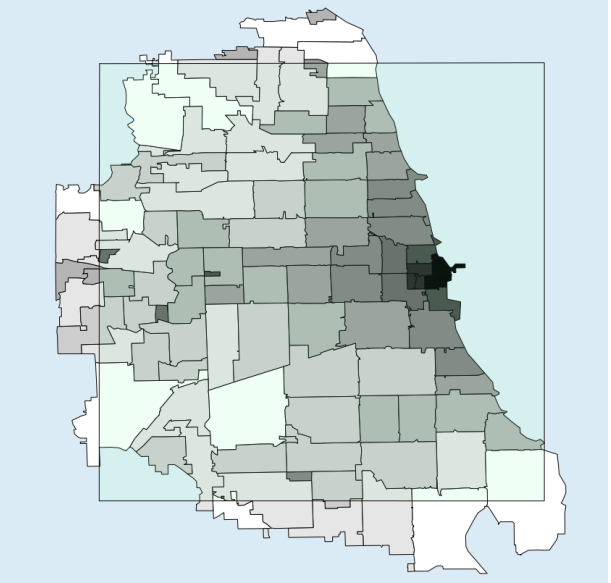
[33]For numerical reasons we impose small moving costs for firms to avoid limit cycles where firms try to arbitrage away tiny profit differences (recall that everything is indivisible in this model). This does not drive our results, is realistic, and introduces some additional randomness into the initial spatial allocation. It also implies that shocks can have small permanent effects even in the absence of agglomeration economies.

Figure A11: Initial equilibrium distribution of prime services and land rents

(a) Initial equilibrium PS distribution



(b) Initial equilibrium land rents



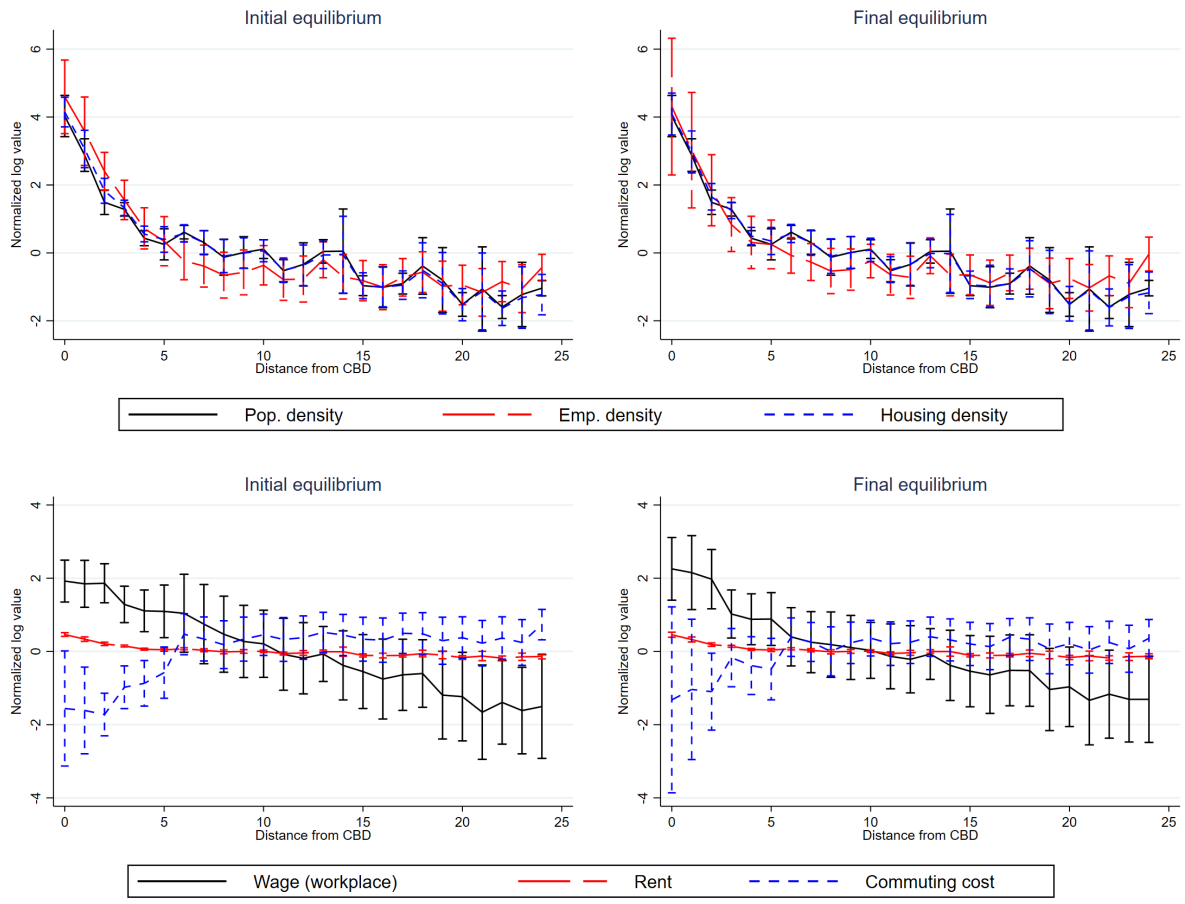
(c) Initial equilibrium rents vs historic land prices in 1913

Notes: Darker colors in panel (a) denote more PS employment; white ZIP Code Tabulation Areas have zero PS employment. In panel (b), darker colors correspond to higher land rents. Panel (c) regresses model land rents (real estate prices) in the initial equilibrium and historic land values in Chicago, 1913 (data from Ahlfeldt and McMillen, 2018) for our simulation runs.

the largest). The initial equilibrium has one prime location that is located in the historic foundation place and that contains 64.3% of the total PS employment in our model city. Observe also that more than half of the locations have zero equilibrium employment. Panel (b) of Figure A11 depicts the corresponding spatial structure of land rents in the initial equilibrium. As shown in panel (c) of Figure A11, this allocation mimics well the qualitative structure of the 1913 Chicago land price gradient.

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,

Figure A12: Various distance gradients in the GSM



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.

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 with 439 employees gets hit by a shock that forces a subset of firms—accounting for 81.5% of the PS employment of the prime location—to move. This leads to a new short-run equilibrium with two locations that account for the lion's share of PS employment totaling 528 and 70 employees, respectively. It is worth noting that the prime location also has moved. Once the shock dissipates and all firms re-optimize, the larger PL in the new location absorbs the smaller one and grows to 719 employees (84.3% of PS 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.

Last, Figure A12 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.

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.

Busetti, 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 an ed. 1986.

Such, Richard and Susan B. Carter, eds, *Historical statistics of the United States—millennial edition online*, Cambridge University Press, 2000.