Capital Cities and Road Network Integration: Evidence from the U.S.

ETH zürich

KOF Swiss Economic Institute

Capital Cities and Road Network Integration: Evidence from the U.S.

Nicole Loumeau

KOF Working Papers, No. 498, November 2021

KOF

KOF

ETH Zurich
KOF Swiss Economic Institute
LEE G 116
Leonhardstrasse 21
8092 Zurich, Switzerland

Phone +41 44 632 42 39
Fax +41 44 632 12 18
www.kof.ethz.ch
kof@kof.ethz.ch

Capital Cities and Road Network Integration: Evidence from the U.S.

Nicole Loumeau*
ETH Zurich

November 22, 2021

Abstract: This paper quantifies the causal effect of capital status on road network integration of U.S. micro/metropolitan statistical areas. Road network integration is defined using a class of measurements that evaluate how well a location is connected to all other locations through the National Highway System (NHS). To tackle the non-random placement of capital cities, I instrument capital status using a k-means clustering algorithm that predicts the boundaries of 48 U.S. states and defines the geographical center as a hypothetical capital location. Overall, I find significant and robust evidence that capital cities are more *directly* integrated in the NHS than non-capital cities of similar characteristics. I discuss two possible mechanisms behind the *capital premium*: (i) the favorable geographical position of capital cities within their state and (ii) a political interest in connecting capital cities well to major urban areas around.

Keywords: capital cities, transport infrastructure, market access, clustering algorithm

JEL classification: O18; R42; R58; C26

* ETH Zurich – KOF Swiss Economic Institute, Leonhardstrasse 21, 8092 Zurich, Switzerland; E-mail: nicole.loumeau@kof.ethz.ch; +41 44 632 33 48.

1 Introduction

A significant share of world trade in goods happens within national borders, via national roads. In 2007, 80 percent of U.S. manufacturing production was traded domestically (Egger et al., 2019) and 44 percent of it using the national road network (U.S. Bureau of Transportation Statistics).[1] Thus, integration into the national transport system is key for any city's economic prosperity. Cities, however, are heterogeneous. One important, yet understudied, dimension in which cities differ is their political status. Investigating whether city heterogeneity, such as political status, affects a city's access to the transport network is important for local economic prosperity, and, therefore, for both researchers and policy makers.

This paper links the political status of U.S. urban areas to their integration in the national road network, in order to understand whether there is a *capital premium* – i.e., a premium to being a state capital city in terms of road infrastructure provision. Road network integration is defined using a class of measurements that evaluate how well a location is connected to other locations through the National Highway System (NHS). The main empirical result suggests that, indeed, U.S. state capitals have on average 14 percent larger levels of (population-weighted) road network integration compared to non-capital cities of similar characteristics.

The paper contributes to the literature in the following two ways. First, this is the first paper that quantifies the causal effect of political status on road network integration and applies the analysis to 920 U.S. Core Based Statistical Areas (CBSAs). The U.S. offer a unique variation to answer the question at hand. There is a large variation of state capital characteristics, which allow me to differentiate political status from size effects. Moreover, it is the only country in the world where such a large number of states, capital cities and urban areas are connected by a common national road network under the same institutional and cultural framework.

The second dimension in which the paper contributes to the literature is its instrumentation strategy. The location choice for most state capitals was closely related to the westward expansion of the U.S. along historical exploration roads, which are correlated with nowadays' transport network (see Duranton and Turner, 2012). To tackle this concern, I construct an instrument which captures the fact that state centrality – *independent of the transport network* – is a key feature of U.S. capital cities.[2] Formally, I employ a k-means clustering algorithm – a concept that is widely applied in Machine Learning literature – that predicts the boundaries of 48 U.S. states and defines their geographical center as a hypothetical capital location. Then, capital status is predicted by the rank in distance to the respective hypothetical capital location.

For each CBSA, I determine the integration in the NHS by four different measures of road network integration: *connectivity* and *market access*, which are based on absolute distances between locations, and *relative connectivity* and *relative market access*, which are based on relative distances between locations. All measures have in common that their value for a given CBSA is the sum over connections to all other places. While connectivity and market access define a connection as inverse (absolute) distance on the network, relative connectivity and relative market access evaluate the distance on the network relative to the great-circle distance. Overall, I find a positive and significant effect of capital status on relative distance measures. This is evidence for a more direct integration of capital cities in the National Highway System. The effect is driven by capital cities in large states and those with an above-median rate of urbanization (defined at the national level).

There are two reasons why U.S. state capitals are expected to be better integrated in the federal road network than comparably large non-capital cities. First, (most) state capitals embody their role of political

[1] Egger et al. (2019) report country-level information on own consumption in total production of manufacturing goods in US dollars. The U.S. Bureau of Transportation Statistics provides information on shipments by travel mode in U.S. ton-miles of freight.

[2] Campante and Do (2014) and Rossitti (2020) also used the state centrality argument for the instrumentation of capital location. However, my instrumentation approach assumes endogenous state borders, whereas Campante and Do (2014) and Rossitti (2020) take the location of state borders as given.

power by being centrally located and easily accessible from other urban centers around them. This makes U.S. capital cities a natural candidate for a direct integration in the transport network, following Christaller's (1933) Central Place Theory (CPT). The transport principle in the CPT suggests that the most efficient and cost minimizing transport network is one that radially connects the most central place in the hierarchy to all other places in the jurisdiction.

Another reason why capital cities may have been favored in the provision of road network infrastructure is that the decision on new road locations is a highly political one. Highway spending, in particular, responds to strong interest groups, 'pork barrel' projects being an obvious example (see Evans, 1994). In the case of the *Interstate Highway System* (as part of the NHS), states were asked to submit proposals for their portion of the federal highway network in response to the recommended national plan (see Baum-Snow, 2007). The final proposal was quite certainly an outcome of inter-governmental lobbying, both from private and public sector interest groups. Of course, whenever the capital city was also the largest economic center, better road infrastructure provision has a straight-forward economic implication. However, for capital cities with little economic relevance this argument does not hold. In those cases, the political status itself could have been the main driver to attracting better access to road infrastructure.

State capital centrality and the political interest to have state capitals well connected to major urban centers are the main mechanisms that motivate the empirical analysis in this paper. In the discussion, I provide further evidence for both mechanisms that underline an existing *capital premium* in direct road network integration.

This paper relates to several strands of the literature. It first relates to the political economy literature that analyzes the importance of political status for local public policies. For example, two recent studies have emphasized the role of capitals for corruption in U.S. states (Campante and Do, 2014) and for sorting of legislators (Rossitti, 2020). There are several other contributions that concentrate on comparative analyses of national capitals in terms of locational policy agendas (see, i.e., Nagel, 2013, for capital cities of federations; Mayer et al., 2017, for capitals that are not the primary economic city of their nations; Rossman, 2018, for newly established capital cities).

The paper also relates to the new economic geography literature that has emphasized the importance of market access in explaining the spatial distribution of economic activity (starting with Krugman, 1991). Apart from theoretical contributions on market access (e.g., Helpman, 1998; Redding and Sturm, 2008), there is a vast empirical literature that focuses on the relationship between access to markets and economic development (e.g., Davis and Weinstein, 2003; Hanson and Xiang, 2004; Redding and Venables, 2004; Hanson, 2005). Most empirical contributions that specifically estimate the importance of transport infrastructure on urban development (e.g., Banerjee et al., 2012; Baum-Snow, 2007; Michaels, 2008; Donaldson, 2018) face an econometric challenge as changes in the transport infrastructure have both direct and indirect (i.e., general equilibrium induced) effects on the observed location. Donaldson and Hornbeck (2016) provide a methodology for measuring the *aggregate* impact of transport infrastructure changes using a reduced-form market access approach that is derived from general equilibrium trade theory. This paper builds on the insights from Donaldson and Hornbeck (2016) in creating the road network integration measures.

When analyzing the link between transport infrastructure and urban development, most studies differentiate cities according to economic characteristics such as city size, productivity or sector composition. Political status as an additional source of heterogeneity, however, has gained only recently more attention. In particular, there are a few contributions that exploit the economic consequences of relocating or constructing national capitals. For instance, Becker et al. (2018) evaluate the impact of a public employment shock on private sector employment due to the relocation of the German capital from Berlin to Bonn and vice versa. Morten and Oliviera (2018) quantify the effect of an exogenous shock in transport infrastructure on trade and migration, succeeding the construction of Brazil's new capital Brasilia. Bai and Jia (2020) exploit the historical variation in changing provincial capitals to analyze the importance of administrative hierarchy on local development in China. They find evidence that Chinese regimes readjusted the transportation network in favor of prefectures that had capital status.

Finally, this paper relates to a new strand of the economic geography literature that has applied algorithmic approaches in instrumentation strategies (i.e., Faber, 2014; Alder and Kondo, 2018; Egger et al., 2020; Arribas-Bel et al., 2019; de Bellefon et al., 2019). These studies typically use mathematical tools, including machine learning algorithms, to replicate key observed institutional features. In particular, they focus on modeling the main determinant that guides the institutional design. For example, transport networks are constructed along the least cost path (Faber, 2014; Alder and Kondo, 2018; Egger et al., 2020) or urban statistical areas unite locations with similar urban context (Arribas-Bel et al., 2019; de Bellefon et al., 2019). As long as the key determinant is well identified, these approaches will provide powerful instruments. The present paper adopts a similar approach and predicts hypothetical capital locations using the k-means clustering algorithm that is based on geographical and topological data. By minimizing the distance from the hypothetical capital to all points within a cluster, the k-means clustering algorithm exploits the fact that state capital cities occupy central locations within their jurisdictions.[3] As compared to the instrumentation strategy in Campante and Do (2014) or Rossitti (2020), this paper additionally accounts for endogenous state borders.

The structure of the paper is as follows. Section 2 provides a historical background on U.S. American state capitals and the U.S. National Highway System that motivates the empirical analysis. Section 3 presents the data and introduces the measurements of road network integration. Section 4 outlines the identification strategy and the instrumental variable design, and presents the main empirical result. Section 5 discusses drivers and possible mechanisms behind the results. Finally, Section 6 concludes.

2 Historical Background

2.1 U.S. American State Capitals

The historical geography of American state capitals is complex and diverse. This section attempts to summarize their spatial and historical evolution in five major patterns, which will inform the empirical analysis of this paper. The summarized facts build heavily on Christian Montès' (2014) comprehensive contribution on *American Capitals: A Historical Geography*. I start with a brief overview of the U.S. settlement and the evolution of urban centers in the 19$^{\text{th}}$ century.

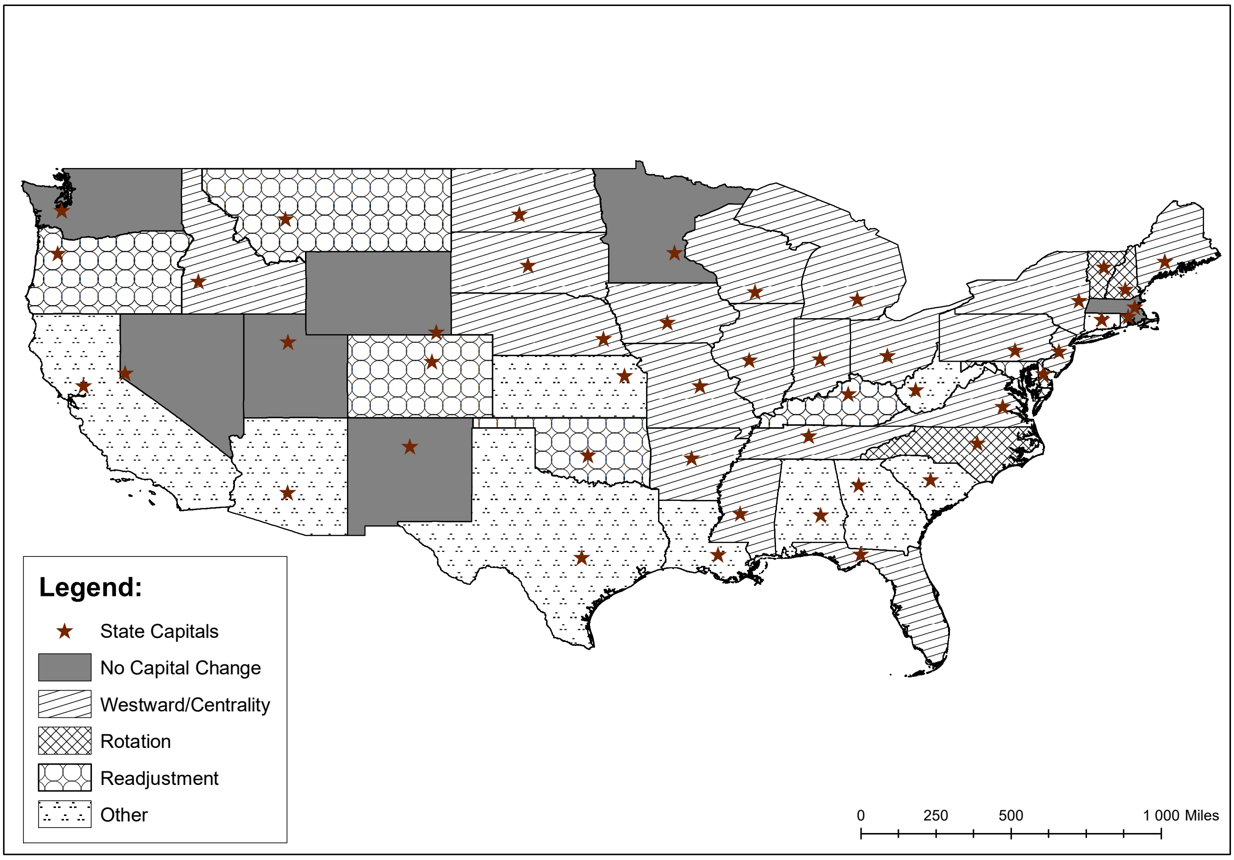
After the creation of the United States of America in 1776, the U.S. territory expanded gradually toward the West, with its first great expansion being the Louisiana Purchase of 1803. The territorial expansion was followed by a substantial shift of population and economic activity from the coast to the center of the territory. U.S. population increased by a factor of eight between 1790 and 1860 and new cities formed, which led to a rapid urbanization during the 19$^{\text{th}}$ century. Two-thirds of the increase in urbanization can be attributed to new cities forming (predominantly) in the South and the Midwest (see Nagy, 2017, for a review on U.S. urban history in the 19$^{\text{th}}$ century).

The land was organized into territories and then states. Once established, states have generally retained their initial borders.[4] Capitals had an important role in territorial structuring as they are places of political power and decision making. Local elites tried to win state capital status not only for economic advantage but also for the political stability that such a status might provide. Most of the time, however, stability did not

[3] A formal analysis of the performance of the algorithm is provided in Section 4.

[4] The exception are four states that have been created from land claimed by another state (Maine, Kentucky, Vermont, West Virginia) and four states (Louisiana, Missouri, Nevada and Pennsylvania) that expanded significantly after acquiring additional federal territory (Zandt, 1976).

Figure 1: Spatial Patterns of Capital Movement



last, and state capitals migrated – on average 3.8 times (Fact 1). The spatial patterns of capital movements were highly heterogeneous (Fact 2). Some states relocated their capitals following the general trends of the U.S. settlements, others experienced a rotation system of capital cities, and again others readjusted their capital location to balance economic and political forces. Eventually, all but eleven of the (present) state capitals were established in the 19$^{\text{th}}$ century, 35 of them before the American Civil War in 1860 (Fact 3). The final decision on capital selection in every state was as heterogeneous as the path that led toward it. However, one striking factor of capital selection was predominant: most states decided against the largest city of the time in the interest of economic and political balance (Fact 4). Some capitals have remained small, others evolved into bustling metropolises. While investigating the reasons for these different development paths might warrant a deeper analysis, it is clearly beyond the scope of this paper. One important aspect that distinguished state capitals at the time, though, is whether or not they were strategically located at important trading routes (Fact 5).

In what follows, I explain the mentioned facts in more detail.

Fact 1: Migration of State Capitals

Most first chosen capitals marked the *entry point* in the *New World* or strategic defense spots that were built to "protect" the pioneers. However, political stability did not last long and capitals migrated – often westward, following the territorial conquest. Only eight states – Hawaii, Massachusetts, Minnesota, Nevada, New Mexico, Utah, Washington and Wyoming – never changed their capital. On average, American states have had 3.84 successive capitals. California, for instance, changed its capital seven times between 1849-1854. How often states have moved their capital does not follow a clear geographical pattern.

Fact 2: Heterogeneous Spatial Patterns of Capital Movements

There are three major spatial patterns of capital movement that amount to 80 percent of all cases.[5]

Westward/Centrality By far the most common reason to relocate the capital (44% of all cases) was due to the *western pull factor* that initiated the movement inland from the coast to the center of a territory. The process first occurred in the East, where coastal capitals had to yield to more centrally located cities, due to the westward expansion inside the state. For instance, New Jersey's first capital Elizabethtown – the port of entry – was relocated 50-miles south-westward and Trenton became the new capital in 1790.

Readjustment Six states (14%) – Alaska, Colorado, Kentucky, Maryland, Montana, Oklahoma, Oregon – relocated their capital in a readjustment process to alter economic and political balance. The readjustment process typically implied a small relocation not far from the first capital choice towards the center of the state.

Rotation Five states (10%) – mostly small ones – experienced a complex system of wandering capitals: Delaware, New Hampshire, North Carolina, Rhode Island and Vermont. This was often an outcome of political rivalries. Apart from Rhode Island, all other states that experienced a capital rotating system chose their permanent capital around the turn to the 19^{th} century. Anecdotal evidence suggests that as the abrupt ending of rotation happened within two decades, the ultimate capital choice was a quasi-random outcome among all geographically central alternatives (Montès, 2014, p.71).

All three major spatial patterns for capital migration are somehow linked to state centrality. Nowadays, U.S. state capitals are on average located in a radius of 70 miles from the state centroid. Figure 1 summarizes the spatial patterns of capital change and shows a map of all states indicating the reason for capital relocation.

Fact 3: Timing of Capital Selection

The majority of state capitals (79%) were selected during the 19^{th} century. With New Mexico being the exception (Santa Fe was chosen already in 1610), the first states to select their permanent capital are – not surprisingly – those along the east coast: Massachusetts (1692), Maryland (1694), Delaware (1781), and Virginia (1779). In total, 35 states had made the decision for their permanent capital before 1860 – just when the American Civil War hit the country and created a large and long-lasting impact on the U.S. economy. With this timing in mind, one could argue that for those 35 states, the reconstruction and subsequent economic development after the Civil War happened with the capital city location as given. Table 8 in the Appendix summarizes the timing of capital selection for all states and adds additional information, including income and population statistics as well as the rank (by population) of each capital city now and then.

Fact 4: Rejection of the Largest City

Even though states and their selection process for capital cities were highly heterogeneous, there is one striking pattern: the majority of states (70%) decided against the largest city (at the time) as their capital. The basic facts in Table 8 in the Appendix compare the population rank of all capital cities at the year of selection to the rank in 2010. Some capitals have remained small (e.g., Frankfort, Kentucky; Annapolis, Maryland; Carson City, Nevada), while others have evolved into the largest cities of their state (e.g., Jackson, Mississippi; Phoenix, Arizona; Atlanta, Georgia). Nowadays, most capitals are de facto large and economically important in their state, though, in 42 percent of the cases they are not *the* largest city.

[5] For more details regarding the spatial patterns of capital movement, see Montès (2014)

Table 1: Descriptive Statistics

	Capitals	Non-Capitals	Top 50 Non-Capitals*
Av. Population (2010)	1,100,626	267,133	2,777,541
Av. Annual Wage (2018)	50,949	45,874	51,556
Av. Area (CBSA)	6,305	2,824	8,113
Av. Population Rank within State	2.75	15.06	2.40
Av. Annual Wage Rank within State	2.16	7.21	4.70
Av. Area within State	3.88	15.00	4.22
Observations	48	872	50

*: by population in 2010. *Data Source:* Wage estimates (2018) for 376 CBSAs are taken from U.S. Bureau of Labor Statistics. Population records in 2010 and area in km^2 in 2017 by CBSA are provided by the U.S. Census Bureau. *Notes:* The average population rank within states ranges between 1-12 for capitals, 1-69 for non-capitals, and 1-9 for top 50 non-capitals. The average wage rank within states ranges between 1-7 for capitals, 1-26 for non-capitals, and 1-24 for top 50 non-capitals. The average area rank within state ranges between 1-18 for capitals, 1-69 for non-capitals, and 1-29 for top 50 non-capitals.

Table 1 presents further descriptive statistics regarding recent population, wages and area for three categories: (present) capital cities, non-capital cities and the top 50 non-capitals cities (by population). It suggests that capital cities are half the size of the largest non-capital cities in terms of population, but four times larger than the average non-capital metropolitan area. Average annual wages in capital cities are similar to wages in the largest non-capital cities, however, capital cities are about 25% smaller in area than large non-capital agglomerations. Within their respective state, capital cities occupy an higher rank in population (i.e., they are relatively smaller in population), but a lower rank in wages and area relative to large non-capital cities.[6]

Fact 5: Capitals as Mercantile Gateways

Capital cities were of primary importance to the developing trading network. Almost all first capitals in the newly settled West were founded at important trading posts (Montès, 2014, p.119). In the 19$^{\text{th}}$ century, inland transportation relied heavily on trails and streams, and consequently, the importance of the road network declined drastically with technological change. However, intra-state and short-distance transportation still depended on the existing road network. In particular, centrally located capitals served as re-distribution hubs to all other populated places within their state.

2.2 The U.S. National Highway System

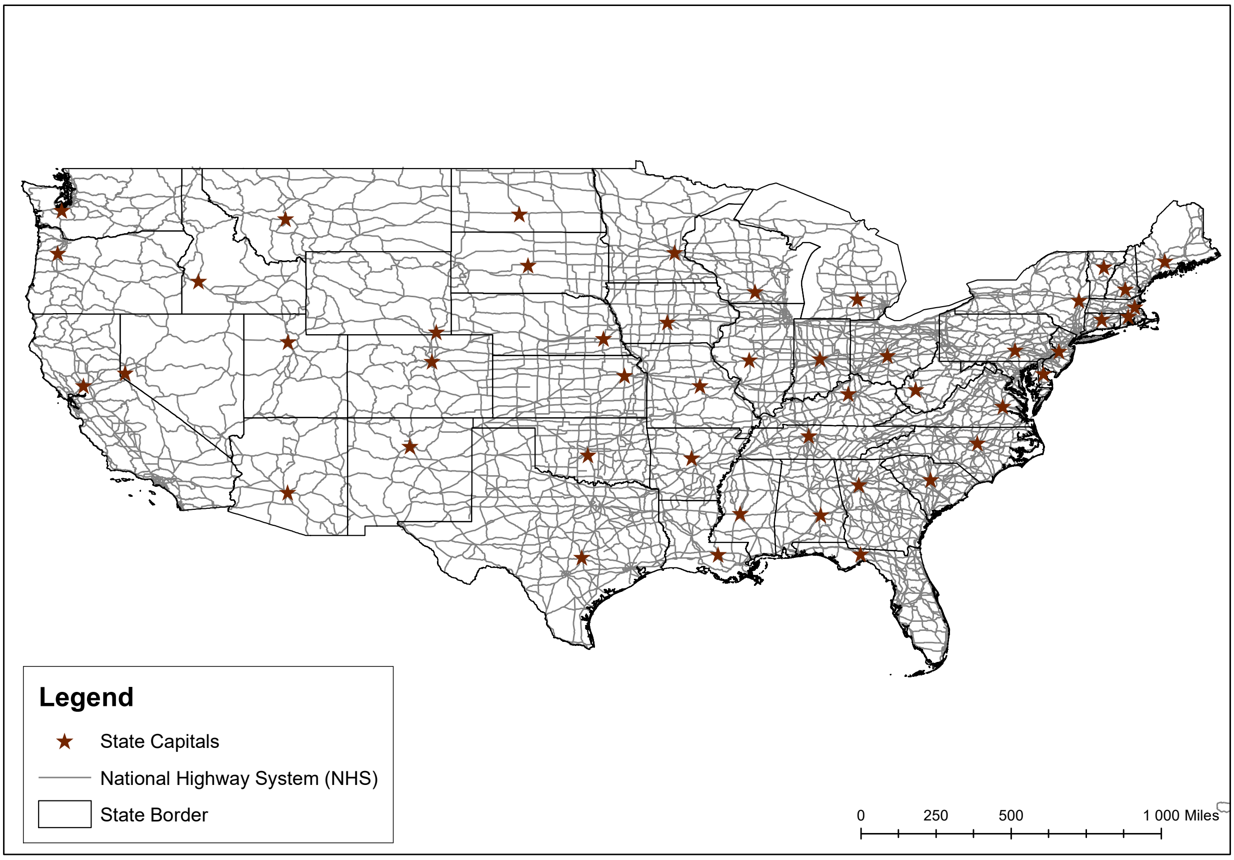
The National Highway System (NHS) constitutes the major federal road network, which strategically connects all states across the U.S. The first federal involvement in developing a national highway system came with the Federal Aid Highway Act of 1944 and the subsequent construction of the Interstate Highway System. By 2011, about 164,000 miles of national highways were completed, of which 47,000 miles compromise the Interstate Highway System. According to the U.S. Department of Transportation, all urban centers with a population of over 50,000 are within five miles of the network.

Figure 2 portrays the National Highway System as of 2017 and highlights the location of state capitals. It shows that the network is more dense in the Northeast – i.e., in proximity of a larger number of high-density urban areas – and less dense in the Midwest and West of the U.S. (except for California).[7] Moreover, the

[6] The wage rank in Table 1 has to be interpreted with caution as data on wage estimates in metropolitan and non-metropolitan areas provided by the U.S. Bureau of Labor Statistics is only available for 376 out of 920 CBSAs.

[7] The larger number of high-density urban areas in the Northeast is also a result of an average smaller state size in north-eastern U.S.

Figure 2: State Capitals and National Highway System



Note: National Highway System as of 2017 (Natural Earth Data Version 4.0.0).

map suggests that in the present network capital cities are well connected. Especially in those states where capitals are centrally located within their state, the federal road network extends radially in all directions, suggesting a direct integration of capital cities (i.e., Arizona, Iowa, Indiana). Additionally, the involvement of local politicians in the design of the national road network may have favored the network integration of capital cities – even in those states where capitals are not centrally located (i.e., Nevada, Oregon).

3 Data Construction

A country like the U.S. is a prime example to analyze the importance of capital status on road network integration for at least two reasons. First, it offers a unique variation of state capital characteristics. This proves particularly useful as it allows me to differentiate political status from size effects. Second, the U.S. is the only country in the world where such a large number of states, capital cities and urban areas are connected by a common national road network under the same institutional and cultural framework. This allows for a uniform measurement of road network integration and the respective covariates and gives enough variation in the empirical analysis.

The units of analysis are Core Based Statistical Areas (CBSA) in the U.S. CBSAs include Metropolitan and Micropolitan Statistical Areas and consist of the county, counties or equivalent entities associated with at least one urban core of at least 10,000 people.[8] After excluding non-contiguous jurisdictions (i.e. Alaska) and off-shore territories (i.e., Hawaii and Puerto Rico) the subsequent analysis includes a total of 920 CBSAs.

Geographical Boundaries and Population The U.S. Census Bureau provides information on geographical boundaries in 2015 and total population estimates between 2010 and 2017 for each CBSA. The geographical extent of a CBSA can be extracted using ArcGIS Software. CBSAs that belong to several U.S. states are attributed to the state in which the main urban center is located (i.e., New York-Newark-Jersey City is assigned to the state of New York even though it extends into New Jersey and Pennsylvania). Historical population records and county boundaries for each decade between 1790-1900 are provided by the National Historical Geographic Information System (NHGIS) Database at the University of Minnesota.[9]

State Capitals The data include a binary indicator for capital status that is unity if a CBSA is a state capital, and zero otherwise. Note that neither the capital of Vermont (Montpelier), nor the capital of Maryland (Annapolis) have an official CBSA definition. I add both to the data and use population levels of 2010 that correspond to the municipal population of Montpelier and Annapolis, respectively.

Road Network Geographical information on the U.S. road network in 2017 is provided by the Natural Earth database.[10] The road network includes major highways, secondary highways, minor roads and ferry routes. In the analysis, I concentrate on major and secondary highways, which broadly define the National Highway System (NHS). Quantifying distances between CBSAs requires a mapping of the geographical division to a single departure or destination point. In the economic geography literature, it is customary to simply assume the centroid of an area. However, given that the connection between CBSAs is of major interest to the analysis in this paper, I create a point measurement that represents the largest concentration of population in a CBSA (i.e., maximum density point) using ArcGIS. In comparison to the centroid of a CBSA, the maximum density point has the advantage that it represents the point from which most people (in expectation) commute or migrate from. This reduces a potential measurement error in establishing the distance between each CBSA pair.

Measuring Road Network Integration

Road network integration is defined using a class of measurements that evaluate how well a location is integrated in the National Highway System. In total, I consider four different measures: connectivity and market access, which are based on absolute distances between locations, and relative connectivity and relative market access, which are based on relative distances between locations. All measures have in common that their value for a given CBSA is the sum over connections to all other locations. While connectivity and market access define a connection as inverse (absolute) distance on the network, relative connectivity and relative market access evaluate the distance on the network relative to the great-circle distance. Hence, relative distances measure network integration in terms of how *directly* a CBSA is connected to all other locations.

In defining the four measures, I denote d_{od} as the shortest distance between an origin (o) and a destination (d) on the road network. Given that most CBSAs are not directly located on the road network, d_{od} is further

[8] Metropolitan Statistical Areas (MSAs) are based on urbanized areas of 50,000 people or more. Micropolitan Statistical Areas (μSAs) are based on urban clusters of at least 10,000 but less than 50,000 people. Adjacent counties become part of a larger urban entity if they have a high degree of social and economic integration with the core as measured by commuting ties.

[9] Historical population records report information of any settlement above 2,500 inhabitants.

[10] Natural Earth is a public domain supported by the North American Cartographic Information Society. I use version 4.0.0. of the database, which got released in 2017.

defined as

$$d_{od} \equiv \varphi(d_{oN_o} + d_{dN_d}) + d_{N_o N_d}, \tag{1}$$

where d_{oN_o} (and d_{dN_d}) indicate the straight-line distance from the maximum density point of a location o (and d) to the road network N_o (and N_d). Distances to the road network are adjusted by a common factor of $\varphi = 1.4$, adding an over-proportional cost if a location is far away from the road network (as in Donaldson and Hornbeck, 2016). The shortest distance between origin and destination on the road network is denoted as $d_{N_o N_d}$.

In the following, I discuss the theoretical foundation as well as the mathematical definition of each measurement.

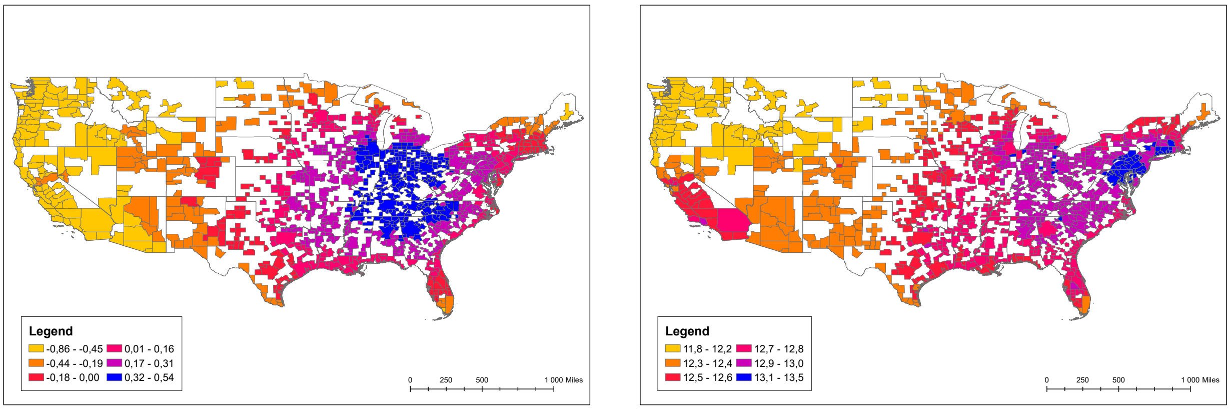
Absolute Distance Measures Absolute distances between locations matter for trade and migration. A first attempt to formally define how well a CBSA is integrated in the road network is to aggregate the inverse of all bilateral distances. Denote connectivity as $Connect_o$, then,

$$Connect_o = \sum_{d \neq o} (1/d_{od}). \tag{2}$$

Higher values of $Connect_o$ indicate smaller aggregate distances to all locations and, hence a better road network integration. A potential concern with the connectivity measure is that it is entirely dependent on the geographical position of a location in space. That is to say, connectivity is naturally larger for those CBSAs that are more centrally located in the road network, as compared to CBSAs at the border of the U.S. territory. To add another (important) dimension to the measure of network integration, one could account for the fact that some connections are economically more valuable than others. For instance, trade theory predicts that proximity to larger markets increases the probability to trade, and hence fosters economic growth. Consequently, considering a measure of market access addresses the economic value of transport connections.

Figure 3: Absolute Distance Measures

Connectivity | Market Access



Note: Connectivity by CBSA is measured as $Connect_o = \sum_{d \neq o} (1/d_{od})$ and reported in logs. Market access by CSBA is measured as $MA_o = \sum_{d \neq o} (L_d/d_{od})$ and reported in logs.

In the new economic geography literature, market access plays a major role in explaining the spatial distribution of economic activity (starting with Krugman, 1991). Donaldson and Hornbeck (2016) derive a first-order approximation for market access from general equilibrium trade theory, which offers an easy application in reduced form analyses. In particular, their approximation defines market access as the sum over the cost of trading with each other location and the other location's population. I follow them and define trade costs as the shortest distance on the network (d_{od}), assuming the elasticity of distance to trade costs to be unity. Then market access, MA_o can be formulated as

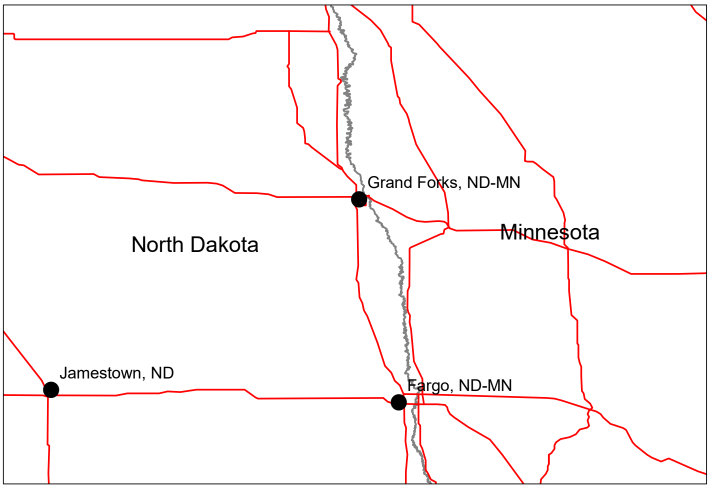
$$MA_o = \sum_{d \neq o} (L_d / d_{od}), \tag{3}$$

where L_d is the population level at destination d.

Figure 3 shows maps of both absolute distance measures at the CBSA level. As expected, connectivity levels are highest in central north-eastern CBSAs and gradually decrease as one moves towards the U.S. national border. The map on market access shows a similar pattern, however, the highest levels are shifted towards highly-populated CBSAs at the north-eastern coast (around Boston and New York). Also, market access levels are high along the south-western coast, due to a large number of high-density places in California.

Relative Distance Measures Contrary to absolute distance measures, relative distance measures evaluate a road network connection relative to the great-circle distance. Essentially, the relative distance captures how *direct* a connection is between a location pair. Figure 4 gives an example of a direct connection and an indirect connection. While the connection from Jamestown to Fargo is almost following a straight line, the connection from Jamestown to Grand Folks requires a detour via Fargo.

Figure 4: Direct vs. Indirect Connection



Notes: The figure portrays the road network between Jamestown (ND), Fargo (ND-MN) and Grand Forks (ND-MN). While Jamestown and Fargo are connected by an almost straight line (direct connection), Jamestown and Grand Forks are connected only via Fargo (indirect connection).

The more a location is directly connected to others the better is its relative connectivity. Formally, relative connectivity is denoted as $Connect_o^R$ and defined as

$$Connect_o^R = \sum_{d \neq o} (d_{od}^{GC} / d_{od}), \tag{4}$$

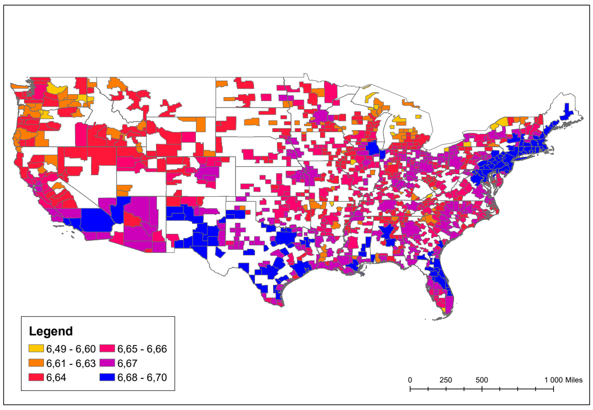
where d_{od}^{GC} describes the great-circle distance between origin o and destination d. The ratio of great-circle distance to network distance, (d_{od}^{GC} / d_{od}), ranges between 0 and 1 for all origin-destination pairs. A ratio close to unity implies that the network distance follows closely the straight-line between the connected places. Weighting the relative connection between two locations by size of the destination market adds an economic value to the relative connection. Hence, the fourth measurement combines market size with the relative distance and formally defines relative market access, MA_o^R, as

$$MA_o^R = \sum_{d \neq o} L_d (d_{od}^{GC} / d_{od}). \tag{5}$$

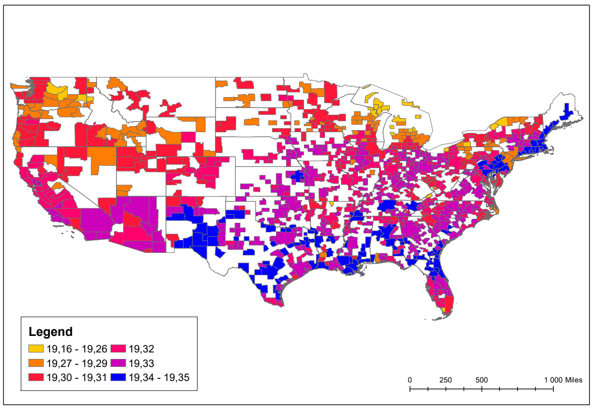
Figure 5 shows maps of relative distance measures at the CBSA level. Two things stand out. First, relative distance measures are less dependent on the geographical position of a CBSA. While there is still a concentration of high levels along the north-eastern coast, relative distance measures show a more significant within-state variation across all U.S. states. Second, both relative connectivity and relative market access clearly identify road network transportation hubs within states. For instance, the centrally located CBSAs in Texas and Alabama are relatively better integrated in the federal road network as those closer to the state border.

Figure 5: Relative Distance Measures

Relative Connectivity



Relative Market Access



Note: Relative connectivity by CBSA is measured as $Connect_o^R = \sum_{d \neq o} (d_{od}^{GC}/d_{od})$ and reported in logs. Relative market access by CSBA is measured as $MA_o^R = \sum_{d \neq o} L_d(d_{od}^{GC}/d_{od})$ and reported in logs.

A potential concern with relative distance measures is that a direct connection between location pairs may be crucially dependent on natural features surrounding them. In other words, connecting CBSAs that are located in, say, the Rocky Mountains may require a large deviation from the great-circle distance solely due to terrain ruggedness. To show that this concern does not systematically affect the relative distance measures, I run a simple OLS estimation that relates relative distance measures with elevation levels and present the results Table 9 in the Appendix. Furthermore, I calculate the average relative connectivity and average relative market access by state and plot the outcome in Figure 6.[11] For a better comparison, I add an elevation heat map in the lower panel of Figure 6.

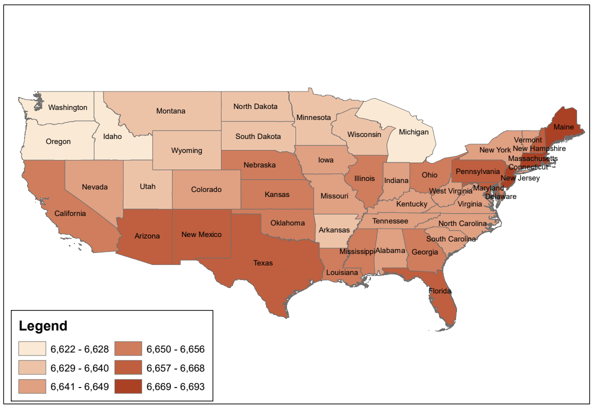
The maps in Figure 6 suggest that terrain ruggedness is not a strong determinant of (average) relative distance measures.[12] Neither states in the Rocky Mountains (e.g., Idaho, Montana, Wyoming, Colorado, New

[11] An alternative approach would be to use Dijkstra's (1959) optimal route (algorithm) instead of the great-circle distance. The advantage of using Dijkstra's algorithm would be that topological features enter as inputs into the design of the optimal route. However, the performance of the algorithm relies heavily on how building costs are specified, which is the topic of a large body of literature in engineering and transport design. The possibility to build tunnels, bridges, bypass segments, etc. typically complicates the choice of the appropriate specification. In comparison, the great-circle distance is a simple and intuitive measure. This modeling choice is further supported by the fact that that terrain ruggedness is indeed *not* a strong determinant of relative distance measures in the present application, as revealed by Figure 6.

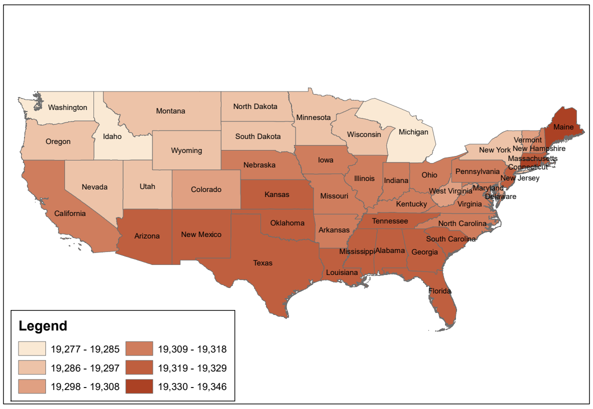
[12] This result is further supported by OLS results in Table 9 in the Appendix. The level of elevation in a CBSA is

Figure 6: Relative Distance Measures (Average by State)

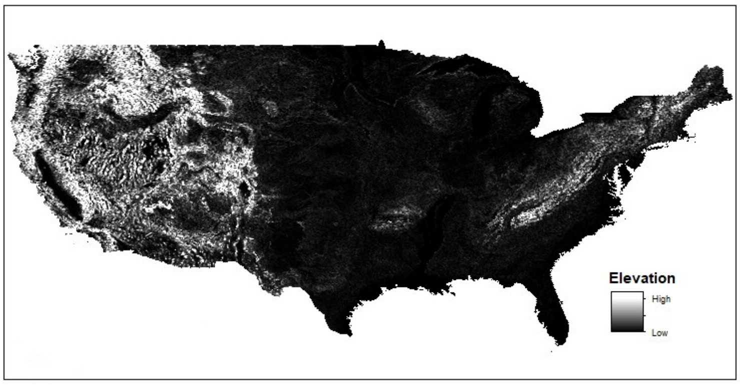
(a) Relative Connectivity



(b) Relative Market Access



(c) Elevation Heat Map



Note: State averages of relative connectivity and relative market access levels are reported in logs, respectively.

Mexico), nor those located along the Appalachian Mountains (e.g., Maine, New Hampshire, Pennsylvania, Virginia) have systematically lower levels in relative distance measures than neighboring states that are not located in the mountains. Instead, there is a clear north-south divide in relative distance measures. To some extent, this captures the higher density of the National Highway System in the Northeast and the South, as compared to the Northwest (compare Figure 2).

4 Empirical Strategy

This section outlines the empirical strategy with a particular emphasis on the Instrumental Variable (IV) design, which allows – given its validity – a causal interpretation of capital status on road network integration.

4.1 Identification

I model the effect of capital status on road network integration using a log-linear specification, that is

$$\log Y_o = \beta \; Capital_o + \boldsymbol{X_o} \gamma + \varepsilon_o, \tag{6}$$

not a significant predictor for relative distance measures, especially, when controlling for additional covariats such as, e.g., population levels and longitude/latitude values.

where Y_o is one of the road network integration outcomes at location o, $Capital_o$ describes a binary indicator that is one if a location is a state capital and zero otherwise, $\boldsymbol{X_o}$ is a vector of covariates of interest, and ε_o is the error term.[13] Retrieving an unbiased estimate for β using Ordinary Least Squares (OLS) requires that capital status was randomly assigned. However, given the historical background of capital selection this assumption would be strong. Even though some capital locations could be defended as (quasi-) random, the location choice for most capitals was closely related to the westward expansion of the U.S. The main concern of endogeneity is that historical routes, such as exploration routes, were a strong determinant for both capital location and today's transport network, which affects in turn the measures of road network integration. Duranton and Turner (2012) provide empirical evidence that historical exploration routes in the U.S. (between 1528-1850) are a strong predictor for nowadays' transport connections. Montès (2014) provides anecdotal evidence that capitals were often located at mercantile gateways (Fact 5) along which the historical road network expanded. Following this argument, OLS would overestimate the true capital effect. On the contrary, there is a possibility that OLS underestimates the true capital effect as most capital cities were not the largest cities in the state (Fact 4) and hence, comparatively less well connected through the (at the time). In any case, the error term ε_o is likely correlated with capital status, $Capital_o$, leading to a biased estimate for β.

4.2 Instrumental Variable Design

To address the endogeneity concern, I construct an instrument for capital status. The instrument exploits the fact that (most) state capitals were chosen for their geographically central and easily accessible location relative to other population clusters in their jurisdiction.[14] I replicate this pattern by employing a heuristic algorithm that predicts the boundaries of 48 U.S. states based on historical U.S. county information and define the geographic center of each predicted U.S. state as the hypothetical capital location.

In Campante and Do (2014) and Rossitti (2020), the location of the state capital is instrumented by the state centroid and therefore follows the same geographical centrality argument as in this paper. However, the main difference here is that U.S. state borders are assumed endogenous, whereas Campante and Do (2014) and Rossitti (2020) take them as given. Historical evidence has shown that U.S. state borders have often been simultaneously chosen with the state capital – at least in those states where the capital location did not (or only marginally) change over time (see Montès, 2014). Consequently, the location of state borders addresses the same endogenity concern as the one for capital location and, thus, needs to be taken into account.

Formally, the construction of the instrument proceeds in three steps.

Step 1: Predicting Geography-based Population Density In order to inform the heuristic algorithm, I predict geography-based population density in 1900 at the county level.[15] I use the population distribution that is determined by geographical features because it addresses a potential endogeneity concern, in which the population distribution of 1900 was partially determined by the location of the historical road network. As geographical features I employ three measures: (i) the average gradient of a county c (G_c), (ii) the distance to the first arrival (F_c), and (iii) the distance to a river (R_c). The average gradient of a county is based on gridded elevation data from the U.S. Geological Survey. It addresses the feasibility of settlement

[13] The vector of covariates includes log population levels in 2010 ($\log L_o^{2010}$), log area size ($\log A_o$), absolute values of longitude and latitude ($|lat_o|, |lon_o|$), and binary indicators for the four large U.S. regions ($Northeast_o$, $South_o$, $Midwest_o$, $West_o$). The mapping of each CBSA to one of the four large U.S. regions follows Caselli and Coleman (2001). For reasons of collinearity, only three out of four U.S. region indicators are included in the estimation.

[14] In the empirical analysis, I abstract from changes in capital city location prior to choosing the permanent capital.

[15] Historical population records are only available at the county level. In some cases historical counties were significantly larger than nowadays CBSAs. To avoid measurement error, I refrain from aggregating the historical county level information to the CBSA level.

given topographical constraints. The distance to first arrival is a (straight-line) distance between the county's centroid and the nearest first arrival point of European settlers on the eastern coast and on the western coast, respectively.[16] It accounts for the gradual evolution of U.S. settlement from the coast to the center of the territory. The distance to the closest river is a (straight-line) distance between a county's centroid and the nearest river based on Natural Earth data. It reflects the importance of population clusters close to good trading opportunities. To predict geography-based population density in 1900, I define the following log-linear specification:

$$\log \bar{L}_c^{1900} = \alpha_{1r} \log G_c + \alpha_{2r} \log F_c + \alpha_{3r} \log R_c + \epsilon_c \quad \text{with} \quad G_c, F_c, R_c > 0, \tag{7}$$

where the subscript r on either coefficient $\{\alpha_{1r}, \alpha_{2r}, \alpha_{3r}\}$ stands for region. It describes one of the four large U.S. regions: the Northeast, the South, the Midwest and the West and thereby follows a well-established regional classification according to the U.S. Census Bureau.[17] I estimate (7) for each region separately by OLS and present the estimation outcome in Table 2. As expected, all geography distance measures are negatively related to population density in 1900. Moreover, all measures are highly relevant for counties located in the Northeast, the South and the Midwest. For counties in the West, only the distance to the first arrival is statistically significant.

Table 2: Geography-based Population Density in 1900

$\log \bar{L}_c^{1900}$	(1) Northeast	(2) South	(3) Midwest	(4) West
$\log G_c$	-0.471***	-0.147***	-0.475***	-0.128
	(0.090)	(0.041)	(0.063)	(0.144)
$\log F_c$	-0.449***	-1.595***	-2.163***	-0.699***
	(0.114)	(0.115)	(0.107)	(0.121)
$\log R_c$	-0.136**	-0.262***	-0.225***	-0.086
	(0.055)	(0.038)	(0.035)	(0.066)
Obs.	243	1253	1023	315
Adj R^2	0.164	0.196	0.277	0.139

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 7: Predicted vs. Observed Population Density

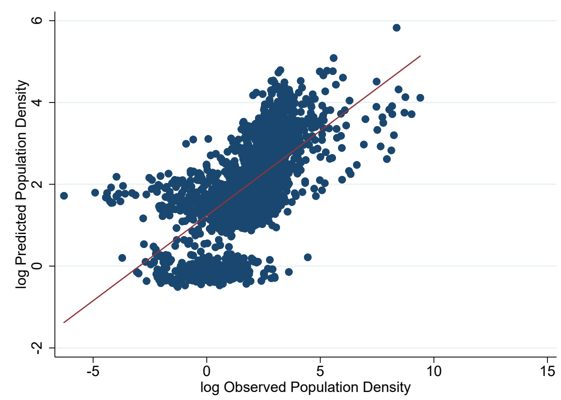


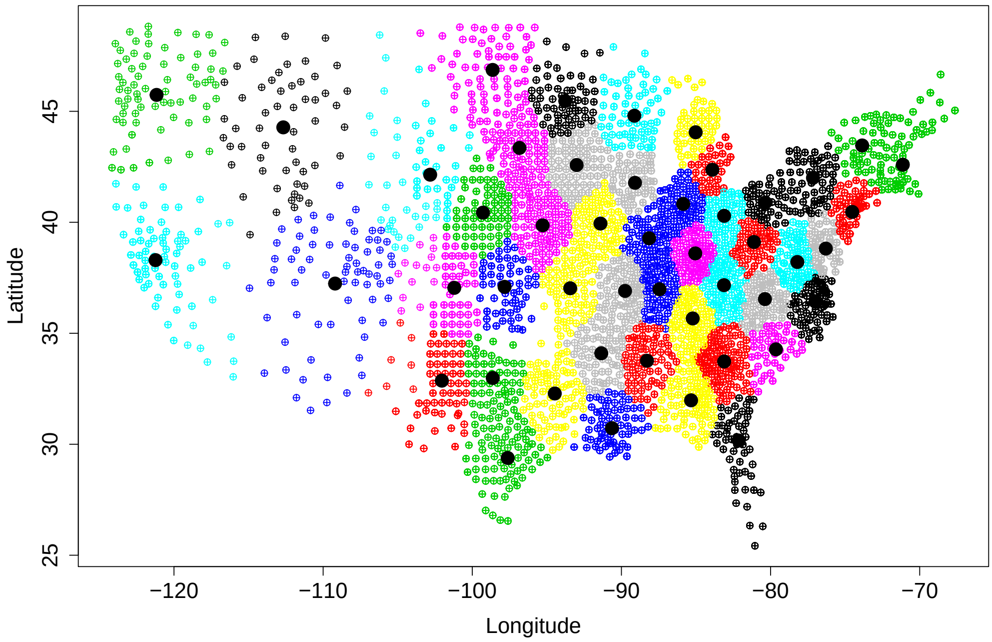
Figure 7 plots the predicted log population density against the observed log population density and highlights the fitted values as a red line. It suggests that the three geography measures replicate a large share of the variation in observed population density. The correlation coefficient between both is 0.64 (p=.000).

Step 2: k-means Clustering Algorithm I use the predicted (log) population density from Step 1 as weights in a k-means clustering algorithm. In my specific application, the k-means clustering provides an answer to the following question: If a central planner had to draw the borders of 48 U.S. states according to the (population-weighted) location of U.S. counties in 1900, where would she draw them? Mathematically, k-means clustering partitions observed counties $c \in C$ into $k = 48$ sets (with $k \leq C$) based on (population-weighted) county coordinate information. The objective is to choose 48 clusters so as to minimize the within-cluster variance. Thereby, I assume the number of clusters as exogenously given, knowing that continental U.S. is

[16]The first arrival point on the east coast is Jamestown, Virginia. The first arrival point on the west coast is San Francisco, California.

[17]I assign each county to one of the four large U.S. regions (the Northeast, the South, the Midwest and the West) following the mapping of Caselli and Coleman (2001).

Figure 8: Predicted U.S. States and Hypothetical Capital Locations



composed of 48 contiguous states.[18] Formally, the algorithm solves the following optimization problem:

$$\operatorname{argmin}_S \sum_{i=1}^{k} \sum_{x_c \in S_i} ||x_c - x_{\mu_i}||^2, \tag{8}$$

where x_c denotes the coordinate point of an observed county c, and x_{μ_i} denotes the k-mean coordinate point of any set $S = \{S_1, S_2, ..., S_k\}$, over which the algorithm optimizes. K-means clustering is a non-deterministic polynomial-time problem, which implies that it is computationally difficult (if not impossible) to determine a global optimum. In the present context, however, reaching a global optimum is not a necessary condition because all that matters is that the instrumental variable meets the validity criteria. By applying an efficient heuristic algorithm the algorithm converges quickly to a local optimum. Ultimately, the hypothetical capital locations are defined as the resulting cluster centers (i.e., k-mean coordinates). Figure 8 shows a cluster plot, which identifies the 48 U.S. states in different colors and marks the hypothetical capital location as the cluster centers in black dots.

Step 3: Construction of the Final Instrument I spatially assign each CBSA in the data to the predicted U.S. state in which the CBSA lies.[19] Once each CBSA is mapped to a predicted U.S. state (k-cluster), I calculate the (straight-line) distance between the CBSA maximum density point and the respective hypothetical capital. I then rank all CBSAs within a predicted U.S. state according to their (straight-line) distance to the hypothetical capital location; and denote this variable as $Rank_o$. In the first stage regression,

[18] By construction, the k-means clustering algorithm improves within-cluster variance as the number of clusters k increases. Taking it to the extreme, the within-cluster variance is *optimal* if each data point is assigned to its own cluster. While this is not a desirable outcome, the data science literature has developed various methods to identify the *appropriate* number of clusters (see Kaufman and Rousseeuw, 1990). The most common methods (i.e., the elbow method or the silhouette method) are based on the idea that adding another cluster is only appropriate if the marginal gain in variance minimization is significantly large enough. In my data, the most appropriate number of clusters according to the silhouette method is $k = 2$.

[19] The exact location of the CBSA is determined by its maximum density point (see Section 3).

$Rank_o$ serves as an instrument that predicts the binary indicator $Capital_o$. Formally, I estimate the following first stage specification:

$$Capital_o = \rho Rank_o + \boldsymbol{X_o}\gamma + \varepsilon_o^{First}. \tag{9}$$

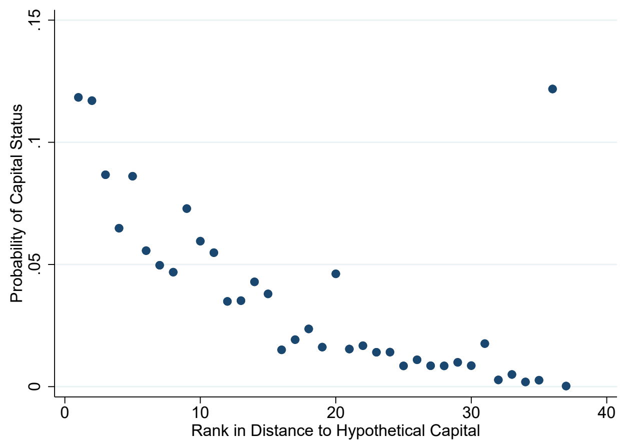
Table 3 presents the first stage results. The instrument is highly relevant and, as expected, it shows a negative sign, which implies that capital cities are in fact nearer located to hypothetical capital locations. The Kleinbergen-Paap F-Statistic for a weak instrument test can be rejected at the 5% significance level.[20] The strength of the instrument is further supported by Figure 9. It plots the probability of being a capital against its rank in distance to the hypothetical capital location and shows a clear negative correlation between the two.

Table 3: First Stage Results

	$Capital_o$
$Rank_o$	-0.003***
	(0.001)
Number of k	48
Observations	920
Adj. R^2	0.12
F-Stat Weak Inst	17.75

Notes: State clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regression includes the list of covariates stated in Section 4.1.

Figure 9: Predicted Capital Status vs. Rank



Note: The graph shows the average probabilty of capital status by rank.

4.3 Results

Table 4 reports the second stage results from estimating (6) for all network integration outcomes. For each outcome, I compare the results of the IV specification to the simple OLS estimate.[21]

When estimating the model with OLS, the effect of capital status on road network integration outcomes is small in magnitude relative to the IV specification. This first finding is surprising. If the capital selection process did, as expected, favor cities that were already well connected, OLS should over-estimate the true capital effect. The small magnitude of the OLS estimates is actually more in line with the historical balance-of-power hypothesis. This hypothesis stipulates that capital status was intentionally attributed to smaller, (initially) less well connected cities in order to spatially separate political and economic centers of power.
Once the endogenous binary indicator $Capital_o$ is instrumented, the effect gets larger in magnitude for all outcomes and highly significant for relative distance measures. Relative distance measures include relative

[20] Stock and Yogo (2005) report critical values at which the weak instrumentation test can be rejected. The critical value is a function of the number of included endogenous regressors, the number of instrumental variables, and the desired maximum bias relative to OLS. In my case, for one endogenous regressor, one instrumental variable and a maximum relative bias of 5%, the critical value is 16.38.

[21] For the sake of brevity, coefficients of all included covariates are suppressed in the main table. The interested reader can find the full table in the Appendix (see Table 10).

Table 4: Second Stage Results

	Absolute Distances				Relative Distances			
	$\log Connect_o$		$\log MA_o$		$\log Connect_o^R$		$\log MA_o^R$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$Capital_o$	0.022	0.298	-0.017	0.470	0.005**	0.151***	0.008***	0.152***
	(0.025)	(0.652)	(0.026)	(0.484)	(0.002)	(0.047)	(0.002)	(0.056)
Observations	920	920	920	920	920	920	920	920
Adj. R^2	0.91	0.61	0.29	0.17	0.87	0.80	0.40	0.34

Notes: State clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the list of covariates stated in Section 4.1.

connectivity and relative market access. The effect of capital status on both is very similar in terms of magnitude and significance. Hence, weighting each relative distance connection by size of the destination market does not change the capital effect overall. In particular, the IV regression result suggests that capital cities have on average about 14 percent larger levels of relative connectivity and relative market access as compared to non-capital cities of similar characteristics.[22] While this result does not have a straight-forward economic interpretation, it confirms that capitals are on average more directly integrated in the road network than non-capital cities.

Absolute distance measures, on the other hand, have a straight-forward economic interpretation. Trade theory suggests that larger levels of connectivity and – even more so of market access – imply better trading opportunities and hence economic prosperity. In the main empirical result, I find a positive though insignificant effect of capital status on connectivity and market access. A possible explanation for why the effect is insignificant could be due to the definition of the absolute distance measures. Both measures are very concentrated in some regions of the U.S. and their magnitude is heavily dependent on how centrally located the CBSA is in the overall National Highway System (see Figure 3). State capitals, however, are naturally very spread out across the entire country. Consequently, the spatial variation in capital city locations is not captured enough by the concentrated measures of connectivity and market access.

5 Discussion

The estimated effect of capital status on road network integration measures is an outcome of a pooled regression, which combines CBSAs of heterogeneous states and capital cities that have been selected for many different reasons. This section sheds further light on the drivers of the effect and provides evidence for plausible mechanisms behind the results. The subsequent analysis concentrates on relative market access as main outcome variable.

Drivers To understand the drivers of the capital effect on relative market access, I construct a number of state-level binary indicators that classify the sample according to general and historical characteristics.

The general characteristics include information on the state urbanization rate, state size and the size of the capital city. Regarding the urbanization rate and state size, I calculate the 50^{th} percentile of the entire distribution and define each binary indicator as unity if a CBSA is located in a state with *below* median urbanization rate and state size, respectively. Regarding size of the capital city, I define the binary indicator

[22] Halvorsen and Palmquist (1980) provide a review on the interpretation of dummy variables in semilogarithmic equations. The effect is calculated as $g = 100 \times (exp(b - V(b)/2) - 1)$, where g is the effect in percent, b is the coefficient on the dummy variable and $V(b)$ is the variance of the coefficient of the dummy variable.

as unity if a CBSA is located in a state in which the capital city is *not* the largest city.

The historical characteristics include information on the spatial patterns of capital migration (see historical background, Section 2). I define four binary indicators for either type of spatial pattern: Westward/Centrality, Rotation, Readjustment and Other.[23] Each indicator is unity if a CBSA is located in a state in which the respective spatial pattern was *not* prevalent.

In separate analyses, I use one binary indicator at the time as an interaction term with capital status, to single out its importance in the overall effect. The empirical specification is as follows

$$\log MA_o^R = \tilde{\beta}\ Capital_o + \delta_1 Indicator_o + \delta_2 Capital_o \times Indicator_o + \boldsymbol{X_o}\gamma + \epsilon_o, \tag{10}$$

where $Indicator_o$ describes one of the previously mentioned binary indicators. Then, the coefficient $\tilde{\beta}$ is the effect of capital status on relative market access conditional on $Indicator_o$ being zero. For this reason, I have defined each indicator in its reverse sense, implying that they are zero for the attribute they are analyzed for.

Table 5: Estimation Results – Drivers

Panel A: General Characteristics				
Dep. Var. $\log(MA_o^R)$	(1)	(2)	(3)	
	Capital is Largest City	**Above Median State Size**	**Above Median Urbanization Rate**	
$Capital_o$ $(\tilde{\beta})$	0.297**	0.454*	0.189***	
	(0.134)	(0.236)	(0.0701)	
Panel B: Historical Characteristics				
Dep. Var. $\log(MA_o^R)$	(1)	(2)	(3)	(4)
	Westward/Centrality	**Rotation**	**Readjustment**	**Other**
$Capital_o$ $(\tilde{\beta})$	0.255**	5.230	2.008	1.105
	(0.102)	(10.98)	(1.761)	(0.864)
Observations	920	920	920	920

Notes: State clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the list of covariates stated in Section 4.1. $Capital_o$ and $Capital_o \times Indicator_o$ are instrumented with $Rank_o$ and $Rank_o \times Indicator_o$.

Table 5 presents the results from estimating (10) as IV regression. The table is divided in two panels and each column in a panel is named after the indicator that is analyzed.[24] When looking at general state characteristics, the results in panel A suggest that the capital effect is driven by large, urbanized states in which the capital is the largest city. Moreover, panel B suggests that the historical decision on state capital centrality within the state is a driver of the capital effect, while other spatial patterns are not.

Centrality Throughout the paper, centrality has played an important role in characterizing U.S. state capitals. By far the most common spatial pattern that decided on the capital location was (geographical) centrality. The employed instrument, that is highly relevant in predicting capital status, is fundamentally based on the idea of (demographic) centrality. The effect of capital status on relative market access is (partially) driven by large U.S. states, where centrality is key to governing the political jurisdiction. In short, centrality matters.

Additional evidence that underlines the centrality argument could be to rerun the analysis with an alternative instrument that is based on geographical centrality within the actual state borders (as in Campante

[23]The category *Other* includes all states that have never changed capital, and those that had other reasons for capital migration than westward/centrality, rotation or readjustment.

[24]For brevity, Table 5 shows only the estimate of interest, $\tilde{\beta}$ (see Table 11 in the Appendix, for full information).

and Do, 2014; Rossitti, 2020). To do so, I rank each CBSA by its distance to the state centroid and denote this instrument as $\widetilde{Rank}_o$. Table 6 contrasts the main estimation result in column (1) to the estimation result with the alternative instrument in column (2). The results suggest that state centrality – based on actual U.S. state borders – is as relevant to predicting the capital location and the capital effect remains positive and significant, though smaller in magnitude.

Table 6: ESTIMATION RESULTS – ALTERNATIVE INSTRUMENT

	First Stage			
		(1)		(2)
Instrument	$Rank_o$	-0.003***	$\widetilde{Rank}_o$	-0.003***
		(0.001)		(0.001)
Dep. Var. $\log(MA_o^R)$	**Second Stage**			
$Capital_o$		0.152***		0.039*
		(0.056)		(0.020)
Observations		920		920
F-Stat Weak Inst		17.75		11.94

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the list of covariates stated in Section 4.1. For one endogenous regressor, one instrumental variable and a maximum relative bias of 5%, the critical value for the weak instrumentation F-Statistic is 16.38 (Stock and Yogo, 2005).

Even though U.S. states and their capital cities are highly heterogeneous, their common feature of state capital centrality is the main mechanism that explains a better – and effectively more direct – road network integration. On the one hand, the Christaller's (1933) Central Place Theory suggests that an efficient road network radially expands around the most central place on top of the hierarchy. On the other hand, even if the capital city is not the largest, most important urban center, its central location favors a better road network integration. This is because, a network that connects places across the entire jurisdiction passes on average more often by the geographic center. Along these lines, Faber (2014) provides evidence that some peripheral places in China have been comparatively well integrated in the National Trunk Highway System due to an *on-the-way* treatment between targeted metropolitan areas.

Connection to Major Urban Centers The second mechanism that may explain the capital effect on (direct) road network integration is related to political interest representation. Conceptually, this mechanism is hard to test for as political influence in road network provision is very difficult to measure. However, there is one plausible argument that interest groups could have defended. As capitals are places of power and decision making, it may be of interest to integrate capital cities well with economically important urban areas around. To test this hypothesis, I construct an alternative measurement of relative market access, $\widetilde{MA}_o^R$, which considers only connections to the 50 largest CBSAs – i.e., those with more than one million inhabitants.
Table 7 compares the main estimation result in column (1) with the effect on the alternative measurement of relative market access in column (2). The comparison shows that the capital effect still holds when only considering connections to the main urban centers, even though the coefficient is slightly smaller in magnitude.

Table 7: ESTIMATION RESULTS – ALTERNATIVE MEASURE OF RELATIVE MARKET ACCESS

	(1) $\log MA_o^R$	(2) $\log \widetilde{MA_o^R}$
$Capital_o$	0.152***	0.147**
	(0.056)	(0.060)
Observations	920	920

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the list of covariates stated in Section 4.1.

6 Conclusion

This paper links the political status of U.S. urban areas to their integration in the National Highway System (NHS) in order to understand whether there is a *capital premium* in road network provision. I document historical patterns of U.S. state capital selection and use the common feature of geographical centrality to construct an instrument for the endogenous capital location. In particular, the IV design is based on a k-means clustering algorithm that predicts the boundaries of 48 U.S. states and defines the geographical center as a hypothetical capital location. I then estimate the causal effect of capital status on four outcomes of road network integration. Two outcome measures (connectivity and market access) evaluate the strength of integration based on the aggregate proximity to all other locations. The other two outcomes (relative connectivity and relative market access) measure how directly connected a location is to all others. I find significant and robust evidence that capital cities are more *directly* integrated in the NHS compared to non-capital cities of similar characteristics. The reason for this finding is a combination of two aspects. First, (most) capital cities have a favorable geographical position within their state. This makes them a natural candidate for a direct road network integration according to the Central Place Theory. And second, as capital cities are places of political power and decision-making, there is a governmental interest in establishing direct connections to other major urban areas. Given that the decision on the location of the federal highway network was subject to inter-governmental negotiations, this interest likely played in favor of capital cities.

References

Alder, S. and I. Kondo (2018): "Political Distortions and Infrastructure Networks in China: A Quantitative Spatial Equilibrium Analysis," 2018 Meeting Papers 1269, Society for Economic Dynamics.

Arribas-Bel, D., M.-A. Garcia-Lopez, and E. Viladecans-Marsal (2019): "Building(s and) cities: Delineating urban areas with a machine learning algorithm," *Journal of Urban Economics*, 103217.

Bai, Y. and R. Jia (2020): "The Economic Consequences of Political Hierarchy: Evidence from Regime Changes in China, AD1000-2000," NBER Working Papers 26652, National Bureau of Economic Research, Inc.

Banerjee, A., E. Duflo, and N. Qian (2012): "On the Road: Access to Transportation Infrastructure and Economic Growth in China," NBER Working Papers 17897, National Bureau of Economic Research, Inc.

Baum-Snow, N. (2007): "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics*, 122, 775–805.

Becker, S. O., S. Heblich, and D. M. Sturm (2018): "The Impact of Public Employment: Evidence from Bonn," CEPR Discussion Papers 12565, C.E.P.R. Discussion Papers.

Campante, F. R. and Q.-A. Do (2014): "Isolated Capital Cities, Accountability, and Corruption: Evidence from US States," *American Economic Review*, 104, 2456–81.

Caselli, F. and W. J. Coleman (2001): "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 109, 584–616.

Christaller, W. (1933): *Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen.*, Jena: Fischer.

Davis, D. R. and D. E. Weinstein (2003): "Market access, economic geography and comparative advantage: an empirical test," *Journal of International Economics*, 59, 1–23.

de Bellefon, M.-P., P.-P. Combes, G. Duranton, L. Gobillon, and C. Gorin (2019): "Delineating urban areas using building density," *Journal of Urban Economics*, 103226.

Dijkstra, E. W. (1959): "A note on two problems in connexion with graphs." *Numerische Mathematik*, 1, 269–271.

Donaldson, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

Donaldson, D. and R. Hornbeck (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 131, 799–858.

Duranton, G. and M. Turner (2012): "Urban Growth and Transportation," *Review of Economic Studies*, 79, 1407–1440.

Egger, P. H., N. Loumeau, and G. Loumeau (2020): "Decomposing the Economic Effects of Transport Infrastructure," Unpublished manuscript.

Egger, P. H., S. Nigai, and N. M. Strecker (2019): "The Taxing Deed of Globalization," *American Economic Review*, 109, 353–390.

Evans, D. (1994): "Policy and Pork: The Use of Pork Barrel Projects to Build Policy Coalitions in the House of Representatives," *American Journal of Political Science*, 38, 894–917.

Faber, B. (2014): "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 81, 1046–1070.

Halvorsen, R. and R. Palmquist (1980): "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 70, 474–75.

Hanson, G. H. (2005): "Market potential, increasing returns and geographic concentration," *Journal of International Economics*, 67, 1–24.

Hanson, G. H. and C. Xiang (2004): "The Home-Market Effect and Bilateral Trade Patterns," *American Economic Review*, 94, 1108–1129.

Helpman, E. (1998): *The Size of Regions*, Cambridge: Cambridge University Press, 33–54.

Kaufman, L. and P. J. Rousseeuw, eds. (1990): *Finding Groups in Data*, John Wiley & Sons, Inc.

Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

Kruskal, J. B. (1956): "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem," in *Proceedings of the American Mathematical Society, 7*.

Mayer, H., F. Sager, D. Kaufmann, and M. Warland (2017): *The Political Economy of Capital Cities*, Routledge.

Michaels, G. (2008): "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 90, 683–701.

Montès, C. (2014): *American Capitals: A Historical Geography (University of Chicago Geography Research Papers Book 247)*, University of Chicago Press.

Morten, M. and J. Oliviera (2018): "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City." Tech. rep., NBER Working Paper 22158.

Nagel, K.-J. (2013): *The Problem of the Capital City. New Research on Federal Capitals and Their Territory*, Colleccio Institut d'Estudis Autonomics, Barcelona.

Nagy, D. K. (2017): "City location and economic development," Unpublished manuscript.

Redding, S. and D. Sturm (2008): "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98, 1766–97.

Redding, S. and A. Venables (2004): "Geography and Export Performance: External Market Access and Internal Supply Capacity," in *Challenges to Globalization: Analyzing the Economics*, National Bureau of Economic Research, Inc, NBER Chapters, 95–130.

Rossitti, G. (2020): "Centres of power US capitals' location and ability sorting of legislators," Unpublished manuscript, London School of Economics.

Rossman, V. (2018): *Capital cities: Varieties and patterns of development and relocation*, London and New York: Routledge.

Stock, J. and M. Yogo (2005): *Testing for Weak Instruments in Linear IV Regression*, New York: Cambridge University Press, 80–108.

Zandt, F. K. V. (1976): *Boundaries of the United States and the several States*, US Geological Survey.

Appendix

7 Supplement Tables

Table 8: State Capitals - Basic Facts

State	**Capital**	**Year chosen**	**Population**	**Population (Census 2010)**				**Average Annual Wage (2018)**		
			Rank when chosen	Rank in 2010	CBSA	State	% State	CBSA	% State	Rank in 2018
Alabama	Montgomery	1847	2	4	374,536	4,779,736	7.84	42,510	97.08	4
Alaska	Juneau	1900	2 or 3	3	31,275	710,231	4.40	N/A		
Arizona	Phoenix	1889	2	1	4,192,887	6,392,017	65.60	50,520	102.50	1
Arkansas	Little Rock	1820	1	1	699,757	2,915,918	24.00	45,020	108.38	2
California	Sacramento	1854	2	5	2,149,127	37,253,956	5.77	56,430	95.40	5
Colorado	Denver	1868	1or 2	1	2,543,482	5,029,196	50.57	59,440	106.49	2
Connecticut	Hartford	1873	1	1	1,212,381	3,574,097	33.92	60,820	100.07	2
Delaware	Dover	1781	1or 2	1	162,310	897,934	18.08	45,120	84.62	1
Florida	Tallahassee	1823	3	12	367,413	18,801,310	1.95	45,340	98.54	7
Georgia	Atlanta	1868	2	1	5,286,728	9,687,653	54.57	52,750	109.26	1
Hawaii	Honolulu	1900	1	1	953,207	1,360,301	70.07	54,870	103.72	
Idaho	Boise City	1864	New Town	1	616,561	1,567,582	39.33	45,470	104.58	1
Illinois	Springfield	1837	2	5	210,170	12,830,632	1.64	50,560	94.00	5
Indiana	Indianapolis	1825	2	1	1,887,877	6,483,802	29.12	49,380	109.03	1
Iowa	Des Moines	1857	7	1	569,633	3,046,355	18.70	52,220	113.15	1
Kansas	Topeka	1861	4	2	233,870	2,853,118	8.20	44,010	97.20	2
Kentucky	Frankfort	1793	2	9	70,706	4,339,367	1.63	N/A		
Louisiana	Baton Rouge	1882	2	2	802,484	4,533,372	17.70	45,200	105.95	1
Maine	Augusta	1832	2	3	122,151	1,328,361	9.20	N/A		
Maryland	Annapolis	1694	N/A	4	38,394*	5,773,552	0.66	N/A		
Massachusetts	Boston	1692	1	1	4,552,402	6,547,629	69.53	67,370	105.41	1
Michigan	Lansing	1847	New Town	3	464,036	9,883,640	4.69	49,320	99.62	5
Minnesota	Minneapolis	1849	1	1	3,348,859	5,303,925	63.14	57,420	105.94	2
Mississippi	Jackson	1826	3	1	567,122	2,967,297	19.11	43,180	109.54	1
Missouri	Jefferson City	1822	New Town	6	149,807	5,988,927	2.50	42,120	90.66	5

Continued.

Table 8: *Continued.*

State	Capital	Year chosen	Population	Population (Census 2010)				Average Annual Wage (2018)		
			Rank when chosen	Rank in 2010	CBSA	State	% State	CBSA	% State	Rank in 2018
Montana	Helena	1874	1 or 2	6	74,801	989,415	7.56	N/A		
Nebraska	Lincoln	1867	New Town	2	302,157	1,826,341	16.54	46,800	100.19	2
Nevada	Carson City	1861	2or 3	3	55,274	2,700,551	2.05	50,840	110.11	1
New Hampshire	Concord	1808	2	3	146,445	1,316,470	11.12	N/A		
New Jersey	Trenton	1790	1	1	366,513	8,791,894	4.17	63,700	109.43	1
New Mexico	Santa Fe	1610	1	3	144,170	2,059,179	7.00	46,260	101.89	2
New York	Albany	1797	2	4	870,716	19,378,102	4.49	54,400	87.93	3
North Carolina	Raleigh	1792	New Town	2	1,130,490	9,535,483	11.86	52,580	111.40	2
North Dakota	Bismarck	1889	4	2	114,778	672,591	17.07	50,390	101.55	1
Ohio	Columbus	1812	New Town	2	1,901,974	11,536,504	16.49	51,260	106.30	1
Oklahoma	Oklahoma City	1910	1	1	1,252,987	3,751,351	33.40	47,120	106.56	1
Oregon	Salem	1860	3	2	390,738	3,831,074	10.20	48,790	93.83	3
Pennsylvania	Harrisburg	1812	5	5	549,475	12,702,379	4.33	49,540	99.02	3
Rhode Island	Providence	1900	1	1	1,600,852	1,052,567	152.09	53,730	98.03	1
South Carolina	Columbia	1790	New Town	2	767,598	4,625,364	16.60	44,680	103.40	2
South Dakota	Pierre	1889	5	9	21,361	814,180	2.62	N/A		
Tennessee	Nashville	1843	1	1	1,670,890	6,346,105	26.33	48,370	108.31	1
Texas	Austin	1839	New Town	4	1,716,289	25,145,561	6.83	53,810	108.23	3
Utah	Salt Lake City	1856	1	1	1,087,873	2,763,885	39.36	50,920	106.26	1
Vermont	Montpelier	1808	6 at most	12	7,855*	625,741	1.26	N/A		
Virginia	Richmond	1779	Village	2	1,208,101	8,001,024	15.10	51,330	92.80	2
Washington	Olympia	1853	Village	4	252,264	6,724,540	3.75	54,050	90.98	4
West Virginia	Charleston	1885	5	2	227,078	1,852,994	12.25	44,680	105.45	2
Wisconsin	Madison	1836	New Town	2	605,435	5,686,986	10.65	52,890	111.70	1
Wyoming	Cheyenne	1869	1	1	91,738	563,626	16.28	48,550	99.84	2

Source: Information on the year of capital choice and rank when chosen are taken from Montès (2014, p.4-5). Population records come from the U.S. Census Bureau, average annual wage estimates come from U.S. Bureau of Labor Statistics. *Notes:* CBSA stands for Core Based Statistical Area. "New Town" means the town was intended to become the capital when founded, "Village" means the town was very small when designated as the capital. *For Annapolis and Montpelier, the population is the municipal population because they are not part of any officially defined CBSA.

Table 9: OLS Results on Relative Distance Measures and Terrain Ruggedness

	Relative Connectivity			Relative Market Access		
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(Elev_o)$	-0.003**	-0.004**	-0.002	-0.004***	-0.002	0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$\log(StateElev_o)$		0.002	0.004		-0.002	0.004
		(0.003)	(0.003)		(0.004)	(0.003)
Additional covariats	NO	NO	YES	NO	NO	YES
Obs.	920	920	920	920	920	920
Adj R^2	0.030	0.032	0.358	0.049	0.053	0.383

Notes: State-clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $\log(Elev_o)$ is log average elevation in a CBSA o. $\log(StateElev_o)$ is log average elevation in the respective state of CBSA o. Data on elevation levels comes from the North America Elevation 1-Kilometer Resolution GRID. Additional covariats include the regressors of interest from the main empirical specification: log population in 2010, log area size, log absolute values of longitude and latitude, binary indicators for the main U.S. regions including, Northeast, the South and the Midwest.

Table 10: Second Stage Results - Full Table

	Absolute Distances				**Relative Distances**			
	$\log Connect_o$		$\log MA_o$		$\log Connect_o^R$		$\log MA_o^R$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
$Capital_o$	0.022	0.298	-0.017	0.470	0.005**	0.151***	0.008***	0.152***
	(0.025)	(0.652)	(0.026)	(0.484)	(0.002)	(0.047)	(0.002)	(0.056)
$\log(L_o^{2010})$	0.027**	0.014	0.079***	0.055*	0.004***	-0.003	0.002***	-0.005*
	(0.011)	(0.035)	(0.014)	(0.030)	(0.001)	(0.002)	(0.001)	(0.003)
$\log(A_o)$	-0.091***	-0.098***	-0.154***	-0.166***	-0.000	-0.004	-0.003*	-0.007*
	(0.019)	(0.026)	(0.019)	(0.027)	(0.002)	(0.003)	(0.001)	(0.003)
$\|lon_o\|$	-0.014***	-0.014***	-0.014***	-0.014***	0.000	0.000	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
$\|lat_o\|$	-0.003	-0.004	-0.011*	-0.013**	-0.002***	-0.003***	-0.003***	-0.003***
	(0.008)	(0.006)	(0.007)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)
$Northeast$	-0.132	-0.142	-0.021	-0.040	0.023**	0.018	0.016*	0.010
	(0.147)	(0.147)	(0.099)	(0.101)	(0.011)	(0.012)	(0.009)	(0.008)
$South$	0.161	0.152*	-0.096	-0.112	-0.002	-0.007	0.005	-0.000
	(0.104)	(0.092)	(0.074)	(0.071)	(0.009)	(0.009)	(0.007)	(0.006)
$Midwest$	0.250***	0.246***	0.007	-0.001	0.008	0.005	0.011**	0.009**
	(0.080)	(0.076)	(0.057)	(0.059)	(0.006)	(0.007)	(0.004)	(0.004)
Cons.	1.712***	1.949***	14.682***	15.102***	6.687***	6.813***	19.378***	19.503***
	(0.534)	(0.498)	(0.381)	(0.332)	(0.032)	(0.052)	(0.031)	(0.046)
Observations	920	920	920	920	920	920	920	920

Notes: State clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $\log(L_o^{2010})$ is log population in 2010, $\log(A_o)$ is log area size, $\log(|lon_o|)$ and $\log(|lat_o|)$ are log absolute values of longitude and latitude, respectively. $Northeast_o$, $South_o$ and $Midwest_o$ are binary indicators that are unity if a CBSA is located in a state that belongs to the Northeast, the South and the Midwest, respectively.

Table 11: Estimation Results - Drivers (Full Table)

Panel A: General Characteristics			
Dep.Var. $\log(MA_o^R)$	(1)	(2)	(3)
	Capital is Largest City	**Above Median State Size**	**Above Median Urbanization Rate**
$Capital_o$ ($\tilde{\beta}$)	0.297**	0.454*	0.189***
	(0.134)	(0.236)	(0.070)
$Indicator_o$	0.018*	0.022**	0.008
	(0.010)	(0.010)	(0.005)
$Capital_o \times Indicator_o$	-0.274**	-0.420*	-0.174***
	(0.129)	(0.226)	(0.067)
$\log(L_o^{2010})$	-0.007*	-0.010	-0.004
	(0.004)	(0.006)	(0.003)
$\log(A_o)$	-0.007**	-0.004	-0.001
	(0.003)	(0.007)	(0.003)
$\lvert lon_o \rvert$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$\lvert lat_o \rvert$	-0.003***	-0.004***	-0.003***
	(0.000)	(0.001)	(0.000)
$Northeast_o$	0.001	0.030**	0.014
	(0.011)	(0.013)	(0.009)
$South_o$	0.000	0.006	0.005
	(0.006)	(0.012)	(0.007)
$Midwest_o$	0.006	0.015	0.013***
	(0.006)	(0.010)	(0.005)
Cons.	19.514***	19.530***	19.452***
	(0.058)	(0.108)	(0.035)

Continued.

Table 10: *Continued.*

Panel B: Historical Characteristics				
Dep.Var. $\log(MA_o^R)$	(1)	(2)	(3)	(4)
	Westward/Centrality	**Rotation**	**Readjustment**	**Other**
$Capital_o$ ($\tilde{\beta}$)	0.255^{**}	5.230	2.008	1.105
	(0.102)	(10.982)	(1.761)	(0.864)
$Indicator_o$	0.015^{***}	0.538	0.127	0.048
	(0.005)	(1.136)	(0.103)	(0.042)
$Capital_o \times Indicator_o$	-0.240^{**}	-5.186	-1.965	-1.065
	(0.098)	(10.920)	(1.739)	(0.842)
$\log(L_o^{2010})$	-0.001	-0.012	-0.015	-0.012
	(0.002)	(0.024)	(0.016)	(0.010)
$\log(A_o)$	-0.010^{**}	-0.003	-0.008	-0.015
	(0.004)	(0.012)	(0.008)	(0.011)
$\lvert lon_o \rvert$	0.000	-0.003	0.000	0.001
	(0.000)	(0.006)	(0.000)	(0.001)
$\lvert lat_o \rvert$	-0.003^{***}	0.000	-0.003^{***}	-0.004^{***}
	(0.000)	(0.006)	(0.001)	(0.001)
$Northeast_o$	0.007	-0.250	0.011	0.035^{*}
	(0.008)	(0.543)	(0.019)	(0.019)
$South_o$	-0.007	-0.020	-0.002	0.009
	(0.006)	(0.051)	(0.012)	(0.016)
$Midwest_o$	0.005	-0.076	0.003	0.014
	(0.005)	(0.170)	(0.014)	(0.012)
Cons.	19.487^{***}	19.258^{***}	19.517^{***}	19.556^{***}
	(0.044)	(0.571)	(0.153)	(0.150)
Observations	920	920	920	920

Notes: State clustered and robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the list of covariates stated in Section 4.1. $Capital_o$ and $Capital_o \times Indicator_o$ are instrumented with $Rank_o$ and $Rank_o \times Indicator_o$. $\log(L_o^{2010})$ is log population in 2010, $\log(A_o)$ is log area size, $|lon_o|$ and $|lat_o|$ are absolute values of longitude and latitude, respectively. $Northeast_o$, $South_o$ and $Midwest_o$ are binary indicators that are unity if a CBSA is located in a state that belongs to the Northeast, the South and the Midwest, respectively.