Identifying Agglomeration Shadows: Long-run Evidence from Ancient Ports

Richard Hornbeck
University of Chicago

Guy Michaels
London School of Economics

Ferdinand Rauch
University of Heidelberg

June 2025

Abstract

We study whether cities create "agglomeration shadows," which discourage economic activity nearby and push it further away. Because identifying such effects is complicated by endogenous city formation, we use data on the locations of ancient ports around the Mediterranean that sometimes seeded cities. We find agglomeration shadows in dense urban activity at intermediate distances from ancient ports, with increased city formation further away. This wave pattern extends to modern city locations more generally and illustrates how encouraging growth in particular places can discourage the growth of nearby areas.

JEL classification: R12, N9
Keywords: agglomeration shadow, urban hierarchy

For helpful comments and suggestions, we thank many colleagues and seminar and conference participants at Bergamo, Berlin, Bern, Bristol, BU, CEU, Census, Dartmouth, FRB-Philadelphia, GEA, Georgetown, Harvard, Humboldt, Konstanz, LSE, NUS, Ottawa, Oxford, PSE, RES, UChicago, UEA, Warwick, Wash U, Yale, and York. In particular, we thank Amir Jina for early conversations, along with Jeffrey Lin and Tomoya Mori. For helpful discussant comments, we thank Arieda Muço and Kohei Takeda. We thank Arthur de Graauw for very helpful advice on using his data on ancient port locations. For excellent research assistance, we thank Sam Abers, Olivia Bordeu Gazmuri, Loughlan O'Doherty, Allen Otterstrom, Georgios Tzortzis, Natalie Yang, and the team at Digital Divide Data. This research was funded in part by the Initiative on Global Markets at the University of Chicago Booth School of Business, the Neubauer Family Faculty Fellowship, and the ESRC's Centre for Economic Performance.

Socrates: We who live between the Pillars of Herakles and Phasis inhabit some small part of it around the sea, just like ants or frogs around a pond.

– Plato (360 B.C.E.), in *Phaedo* 109a-b

Cities benefit from agglomeration spillovers that increase productivity, but may also cast "agglomeration shadows" that discourage some economic activities in surrounding areas and allow other cities to thrive only further away. As dense central cities grow and generate substantial wealth, classic models predict a monotonic decline in economic activity with distance to the central business district (von Thünen, 1826; Mills, 1967; Muth, 1969). In modern quantitative spatial models, better access to markets is generally unambiguously good. With agglomeration shadows, however, the impacts of market access become non-monotonic and nearby areas become relegated to second-tier status in an urban hierarchy with increased economic activity only at further distances (Fujita, Krugman, and Mori, 1999). Agglomeration shadows cast by cities generate long-lasting spatial path dependence from early economic advantages in particular locations. Given sizeable urban-rural differences in productivity, wages, and land values, spatial inequality is amplified when city growth in particular areas discourages city growth in nearby areas.

Longstanding interest in the spatial distribution of economic activity, around the Mediterranean in particular, dates back at least to Plato's *Phaedo* from 360 B.C.E. Early models of urban hierarchies feature larger cities spaced out and interspersed with smaller cities (Christaller, 1933; Jefferson, 1939; Lösch, 1940). The New Economic Geography (NEG) literature modeled how such a system of cities can emerge in theory (Fujita and Krugman, 1995; Fujita, Krugman, and Mori, 1999; Fujita, Krugman, and Venables, 1999). Fujita, Krugman, and Mori (1999) start with a central city, in which firms cluster to benefit from agglomeration forces and increasing returns to scale. As the aggregate population grows, new cities emerge. Although some small cities may emerge closer to the central city, larger cities only develop further away because they face strong competition from the central city without benefiting from its agglomeration forces. The central city thereby casts a shadow, discouraging some economic activity nearby. There is limited causal evidence, however, on whether cities systematically crowd out nearby economic activity that can only thrive further away (as hypothesized by Krugman (1995)).

There are two main reasons why identifying agglomeration shadows empirically has proved difficult. First, cities grow in particular places endogenously in ways that both reflect and influence the suitability of nearby locations for cities. In empirical settings, there is generally not an exogenously-fixed starting central city from which to consider the causal impacts on surrounding areas. Second, as we show in simulations of a canonical NEG model from Fujita, Krugman, and Mori (1999), a form of "wave interference" emerges when estimating average impacts on the surroundings of central city locations. This wave interference arises because small differences across urban networks make their spatial wave patterns asynchronous, which obscures the alternating empty spaces and cities (especially for smaller cities that are often not spaced as far apart).

To overcome endogeneity in city locations, we study shadows cast by cities "seeded" at 4,263 ancient port locations that surrounded the Mediterranean, Black Sea, Red Sea, and nearby coastal areas. We find that ancient port locations are more likely to have grown modern cities than geographically similar locations. Where a natural harbor remains at that ancient port location, the modern city is often located precisely at that location. But where the ancient port location no longer has a natural harbor, a city is often found near that port's location but not at it, and this slight drift over a few kilometers exacerbates the wave interference.

We then estimate the probability that modern cities emerged at various distances from the ancient port locations, reflecting the agglomeration shadows cast by cities seeded at those ancient port locations. Agglomeration shadows generate non-monotonic patterns in economic activity, characterized not only by an absence of other cities nearby but also by an increased likelihood of city formation at a further distance. Ancient ports may have direct effects on modern activity at the port location and its immediate vicinity, but our identification assumes that effects at further distances only reflect agglomeration shadows from cities seeded around the ancient port locations.

The empirical analysis uses population data for 1km-by-1km grid cells (CIESIN, 2018), which cover 2.3 million square km within 200km of an ancient port and within 50km of the coast. We examine the spatial distribution of population density in grid cells around ancient port locations, controlling for grid cells' distance to the coast and nearest river, as well as other "first nature" geographic features that may affect city formation (e.g., ruggedness, temperature, precipitation). We also report estimates that control for endogenous features of the ancient era, such as distance to Roman roads or proximity to ancient cities. The statistical inference we use is robust to adjustments for spatial correlation across grid cells (Conley, 1999; Colella et al., 2019).

We first estimate that grid cells with ancient ports have higher modern population density, on average, and are more likely to have become urban areas or cities than geographically similar grid cells. Grid cells with ancient ports are 12 percentage points more likely to have urban population density (ln density > 6) and 1.7 percentage points more likely to have city population density (ln density > 9), whereas 7.1% of sample grid cells have urban density and 0.2% have city density. These estimates are effectively a "first stage" for our main analysis of impacts on areas further from ancient port locations.

We estimate that average population density declines monotonically in distance from the ancient port locations. The impacts are similar – smaller, but still substantial – around ancient port locations that have since lost their natural harbor. These estimates suggest path dependence in economic activity, not only in particular locations with obsolete geographic advantages, but also in ways that extend to surrounding locations.[1]

We then explore the existence of "agglomeration shadows" around these ancient port locations, which are detectable in our simulations at higher thresholds of population density. We find that urban activity (ln density > 6) declines monotonically in distance from the ancient ports with harbors, whereas city activity (ln density > 9) declines more quickly out to 20km and then increases by 40km. This emergence of cities beyond the agglomeration shadow is smaller than the direct path dependence impact in ancient port locations, but the upward wave in city formation is large relative to the overall low probability of city formation in an average grid cell. The estimated agglomeration shadow in city formation at 10-30km from ancient port locations is similar to the distance of a typical day's travel for pack animals or carts in the ancient world (Scheidel, 2015) when the urban hierarchy was emerging.

Consistent with our simulated predictions for agglomeration shadows, the likelihood of finding city density follows a distinctive wave pattern around the locations of ancient ports whose natural harbors survived and pinned down the location of their seeded city. Agglomeration shadows are more obscured around ancient ports that lost their harbors, for which there is more wave interference. In contrast to the wave pattern for city densities, we observe monotonic declines in urban activity that are similar to classic models of monocentric cities (Alonso, 1964; Mills, 1967; Muth, 1969), but also consistent with "wave interference" from a combination of positive and negative spatial impacts at different distances.

The agglomeration shadows in city activity reflect general city-to-city crowd-out, rather than direct competition between ports themselves. We estimate that ancient port locations are not more likely to have modern port structures when nearby ancient port locations have lost their harbors (and thereby have less port activity). By contrast, these ancient port locations are in competition for more general economic activity: ancient port locations have higher population density, urban activity, and city activity when nearby ancient port locations have lost their harbors (and thereby have less general economic activity).

[1] We focus on the presence of a natural harbor in modern satellite images, setting aside the existence of modern human-made harbors, which are potentially endogenous to local economic demand. Our estimates are also not sensitive to omitting a small number of cases with known ancient investments in open-water ports.

While all cities have a boundary, just beyond which there is no city, the appearance of "agglomeration shadows" is not mechanical. Regions may well contain multiple cities, spaced out, without cities actively pushing other cities away and depressing nearby economic activity. To complement our main analysis, we also develop an alternative empirical approach that uses data on all modern cities in our sample region. Examining the spatial distribution of all modern cities, we find indications of agglomeration shadows relative to a simulated random benchmark. We calculate the distance from each modern city to the nearest modern city (from GHS-UCDB 2019 data), and we plot the cumulative distribution function (CDF) of distances. We compare this observed CDF to the distribution of CDFs for minimum distances between randomly located cities, where these random locations are selected in simulations using probability weights to reflect geographic characteristics that predict city formation. Consistent with the existence of agglomeration shadows in the modern spatial equilibrium, more generally, we observe fewer large cities ($>$ 500,000 people) whose nearest large city is within 40km than in the random simulations. We also observe more large cities from 40km to 60km than in the simulations.

Our main contribution is to use the historical influence of ancient port locations to demonstrate causal negative "agglomeration shadows" that are consistent with long-hypothesized theoretical forces (Christaller, 1933; Lösch, 1940; Krugman, 1995; Fujita, Krugman, and Mori, 1999). This echoes the use of establishment openings to demonstrate causal positive "agglomeration spillovers" (Greenstone, Hornbeck, and Moretti, 2010; Bloom et al., 2019; Giroud et al., 2021), which also followed a long intellectual tradition (Marshall, 1890; Ellison and Glaeser, 1997; Glaeser and Gottlieb, 2009). Our estimates draw on a literature documenting local path dependence (Bleakley and Lin, 2012; Michaels and Rauch, 2018; Allen and Donaldson, 2022), and we extend this literature to consider path dependence in spatial relationships beyond those locations. We show that these resulting agglomeration shadows can be obscured by a pattern of "wave interference." which we document using simulations of the model in Fujita, Krugman, and Mori (1999). Identifying agglomeration shadows is also complicated by classic challenges to causal inferences from endogenous city formation. The locations of ancient ports, and variation in the survival of natural harbors, provide empirical traction to estimate the presence of agglomeration shadows.

Agglomeration shadows reflect more than the tension between agglomeration and congestion forces. Multiple cities form, balancing agglomeration and congestion forces along with geographic variation in natural features (Fujita, Krugman, and Venables, 1999). City systems can arise without systematic shadows (Henderson, 1974), and there could even be advantages to cities forming near other cities. Agglomeration shadows reflect how cities themselves create a distinctive nonrandom spacing, with early successful locations pushing away other cities while encouraging cities to form just beyond that shadow.

These can be subtle forces to detect, though, requiring regions with sufficient economic activity for multiple cities to form and enough homogeneity in economic activity and transportation costs for these shadows to align across areas. Indeed, agglomeration shadows are clearest in our sample within the Roman Empire, and undetectable outside the Roman Empire where there were fewer cities and more heterogeneity in transport costs.

The existence of agglomeration shadows implies broader impacts of place-based policies. Similar types of agglomeration shadows may also arise in access to specific services at different locations, such as spatial differences in healthcare access (Finkelstein, Gentzkow, and Williams, 2021; Dingel et al., 2023) or food deserts (Cummins and Macintyre, 2002). We focus on identifying how major cities affect city formation in surrounding areas, using the locations of ancient ports to gain empirical traction in identifying forces long hypothesized but relatively overlooked in the recent focus on identifying positive agglomeration spillovers. The empirical identification of agglomeration spillovers has helped renew attention to industrial policy, whereas the identification of agglomeration shadows highlights how local successes can discourage economic activities in surrounding areas and amplify spatial inequality in land values and wages.

Section I discusses related literature, simulations of agglomeration shadows in the canonical "new economic geography" (NEG) model of Fujita, Krugman, and Mori (1999), and implications for empirical analysis of agglomeration shadows using ancient ports. Section II introduces the data, with more details in Appendix A, and Section III describes our estimating equations. Section IV reports our estimates using ancient port locations, and Section V analyzes agglomeration shadows among modern cities more generally relative to a random benchmark. Section VI concludes.

I Related Literature and Model Simulations

I.A Agglomeration Spillovers and Agglomeration Shadows

Economic activity is spatially concentrated, partly because of natural advantages to particular locations but also because proximity to other economic activity increases productivity due to "agglomeration spillovers." Manufacturing industries, for example, are more spatially concentrated than can be explained by natural advantages or random chance (Ellison and Glaeser, 1997, 1999; Duranton and Overman, 2005; Ellison, Glaeser, and Kerr, 2010; Mori and Smith, 2015). Indeed, the opening of large manufacturing establishments increases the productivity of nearby incumbent establishments (Greenstone, Hornbeck, and Moretti, 2010). These direct productivity benefits do not extend to further geographic areas but do spread to further establishments owned by those incumbent firms (Giroud et al., 2021), in part through changes in management practices (Bloom et al., 2019).

Cities are not just points in space, representing central business districts where these positive agglomeration spillovers may be strongest. In classic models of "monocentric cities," population density and land prices are highest at the city center and decrease in distance from this center (von Thünen, 1826; Alonso, 1964; Mills, 1967; Muth, 1969; Duranton and Puga, 2004). This empirical pattern is common, though there are exceptions and extended models consider "polycentric cities" (Ahlfeldt and Wendland, 2013; Ahlfeldt et al., 2015). The broader economy then includes a distribution of cities of different sizes (Henderson, 1974; Gabaix, 1999), with a greater range of economic activities in larger cities (Davis and Dingel, 2020).

Strong agglomeration forces raise the prospect of "agglomeration shadows," whereby concentrated economic activity in cities actively discourages some nearby economic activities. In early models of urban hierarchies (Christaller, 1933; Lösch, 1940), a featureless plain is evenly populated by farmers and is served by a hexagonal lattice of "central places" in a hierarchy: larger cities are spaced far apart, producing a broad range of urban goods and services, and smaller cities fill in intermediate areas to offer a narrower range of goods and services.

The tension between congestion and dispersion forces helps shape the spatial distribution of economic activity (Duranton and Puga, 2023). Models in the New Economic Geography (NEG) literature provide micro-foundations that rationalize the emergence of urban hierarchies through a decentralized process (e.g., Fujita and Krugman 1995; Fujita, Krugman, and Mori 1999; Fujita, Krugman, and Venables 1999; Fujita and Mori 2005; Mori et al. 2023).[2] These models predict that cities create agglomeration shadows, discouraging city formation in surrounding areas, but the length of these shadows varies by cities' sizes and the diversity of their economic activity. Large cities are less likely to form near other large cities in these models, but smaller cities may still form closer to large cities.

I.B Agglomeration Shadow Appearance in Simulations

We use a canonical NEG model from Fujita, Krugman, and Mori (1999), "FKM model," to simulate how agglomeration shadows might appear in data when estimating average impacts across geographic regions. Our discussion focuses on bringing out the empirical implications of the model, deferring further theoretical details to Fujita, Krugman, and Mori (1999).

The FKM model describes a one-dimensional economy, with a fixed initial city and other locations indexed by distance r from that initial city. There are two sectors in the model, "manufacturing" and "agriculture." Within manufacturing, there are three industries that

[2] See also a literature on regional variation in industry activity from the Home Market Effect (Helpman and Krugman, 1985; Matsuyama, 2017; Costinot et al., 2019).

each produce a continuum of differentiated goods using labor and production technology with increasing returns to scale. Manufacturing concentrates in cities, whose number and location are determined endogenously. The agricultural sector produces one homogeneous good using labor and land in rural areas, with constant returns to scale production.[3] Goods are traded with iceberg transportation costs. Consumers have Cobb-Douglas preferences over agricultural goods and composite indices of manufactured goods in each industry, with constant elasticity of substitution (CES) preferences over varieties of manufactured goods. Within manufacturing, industry 1 has the highest elasticity of substitution across varieties, followed by industry 2 and then industry 3, so that consumers are most willing to incur transportation costs for industry 3 goods, followed by those of industry 2. Labor is supplied by consumers, who are fully mobile across locations and sectors, and land quality is constant.

When aggregate population N is small, all manufacturing occurs in the fixed initial city. As N increases, the agricultural frontier expands. Once N increases sufficiently, there is enough market potential to produce some manufactured goods away from the central city, and a "third-order" city emerges with manufacturing production only in industry 1. As N grows further, the agricultural frontier keeps expanding and additional cities emerge: more single-industry "third-order" cities, and a small number of two-industry "second-order" cities that are further apart from the central city and each other. The initial city is the only "first-order" city, by assumption, with all three manufacturing industries.

In simulations, we explore how this model generates average impacts on city formation around fixed initial starting locations in independent economies of different sizes. We assume a range of values for aggregate population N that follows FKM, and we graph the share of cases in which cities form at each distance r.[4] The model is symmetric around zero, and like FKM we show locations that are on the positive side, distance r from the fixed starting location.

Figure 1 shows the average spatial distribution of cities. There is always a city at the origin, by assumption, and some distances often have a city, while others rarely do (Panel A). Panel B shows the averaged spatial distribution of "large cities," which are second-order or first-order cities with two or three manufacturing industries. These second-order and first-order cities have larger populations in the model, on average, although population density is not well defined in the model, because cities do not take up space.

[3] "Manufacturing" and "agriculture" could reflect a range of goods and services, where the important feature of the model is that "manufacturing" is done in cities with increasing returns to scale and "agriculture" is done across places with constant returns to scale.

[4] We use the 16 distinct values for aggregate population N reported in Figure 7 of Fujita, Krugman, and Mori (1999). Where there are bifurcations, we choose the side of the bifurcation associated with higher population.

For each aggregate population level, the FKM model determines city locations relative to the fixed starting point where the central city emerges. We also consider the possibility that the central city may not emerge exactly at a fixed point (e.g., an ancient port). Instead, we allow the initial city to form near the fixed starting point, at a location drawn randomly from a normal distribution centered on the fixed starting point with a σ standard deviation. We simulate the model 2,000 times, for each level of aggregate population, and graph the share of cases in which cities form at each distance r, only on the positive side of the fixed starting point.

Figure 2 shows the share of cases in which cities form in each location, with increasing spatial noise. There are visible waves in the probability of any city when allowing for small spatial noise (Panel A), and longer waves in large city locations (Panel B). Allowing for more spatial noise, the peaks and valleys in small city locations overlap and this "wave interference" generates a monotonic decline in the frequency of any city in distance from the origin (Panel C).

This notion of "wave interference" relates to asynchronous waves offsetting each other, which here arises from high-frequency probability density functions for city locations that obscure each other's signal. The locations of large cities are less subject to this wave interference because of their lower frequency, which corresponds to larger gaps. Thus, as we move away from the origin, the probability of finding a large city decreases and then increases (Panel D), reflecting the still-detectable shadow cast by the largest city. But as spatial noise increases further, there is a near-monotonic decline for both all cities and large cities (Panels E and F).

Figure 3 overlays the spatial distributions of cities, with more or less spatial noise, for all cities (Panel A) and large cities (Panel B). Panels C and D report the differences in these lines, with a dip and rise in large city formation probability (Panel D) and a smoother decline and leveling out in all city formation (Panel C). Appendix Figures A.1, A.2, and A.3 report similar figures for the probability that particular distances exceed certain population density thresholds, separating distances into equally sized grid cells and normalizing the density thresholds based on the grid cell size and aggregate population.

These simulations suggest how agglomeration shadows appear when averaging impacts on nearby areas from different starting cities, each "seeded" at separate locations (e.g., ancient ports). Agglomeration shadows are most clearly visible for large densely-populated cities when seeds pin down the precise location of starting cities. Due to wave interference, these shadows become less visible as spatial noise in the starting city location increases, particularly shadows for smaller urban areas that are closer together. We then expect agglomeration shadows to be more visible for large cities than smaller cities and when starting city locations are precisely pinned down (e.g., when the natural harbor survived) than when there is more spatial noise in starting city locations (e.g., when the natural harbor disappeared).

We expect a similar form of wave interference from spatial noise due to heterogeneity in the model parameters, which presumably vary across time periods and regions where urban hierarchies emerge. For example, higher or lower transportation costs would shrink or expand the wave pattern in city locations and create wave interference when averaging across those areas. We later consider estimates within the Roman Empire, with more homogeneity in transportation costs and economic features, compared to estimates from areas outside the Roman Empire with more underlying heterogeneity.

Apart from wave interference, there are two other reasons that agglomeration shadows may not be visible. First, the "no black hole condition" may not be satisfied (Fujita, Krugman, and Venables, 1999), which occurs when agglomeration forces are sufficiently strong to outweigh congestion forces and all the manufacturing activity concentrates in one location (e.g., when substitution elasticities are sufficiently low in the FKM model). Although we know this is not the case globally, it may occur in some regional economies. Second, the regional population may be too low for there to be more than one city (particularly more than one large city). This also motivates our exploration of heterogeneity, where population and economic activity were high within the Roman Empire and shadows may be more detectable.

The FKM model provides one underlying micro structure that generates agglomeration shadows, but different micro structures could also generate the empirical occurrence of agglomeration shadows and wave interference obscuring small gaps between smaller cities. FKM focuses on local increasing returns to scale in manufacturing production, which generate agglomeration shadows because nearby (shadowed) places can import goods and do not benefit from local returns to scale. A similar mechanism can also apply in the case of local knowledge spillovers in manufacturing.

I.C Estimating Agglomeration Shadows

The FKM model illustrates endogenous city formation, but takes an initial starting city as given, which highlights a central empirical challenge: finding exogenous "fixed" starting locations for local urban networks. A city may form in a particular place when surrounding areas are unsuitable, rather than the city itself discouraging nearby city formation. Existing empirical papers take city locations as given and estimate how population growth varies with distance to cities (Ali et al., 2009; Tervo, 2010; Bosker and Buringh, 2017; Cuberes, Desmet, and Rappaport, 2021; Beltrán Tapia, Díez-Minguela, and Martinez-Galarraga, 2021). Our main departures from this literature are to use historical determinants of starting city locations for identification, along with connecting our empirical analysis more closely to theoretical predictions from the NEG theory.

We look to estimate how an urban hierarchy emerges, in which cities affect each other's development over a long time horizon. We use ancient ports, which encouraged city growth in particular places due to highly-local coastline characteristics, to estimate impacts on surrounding areas. This is an advantage of ancient ports, in contrast to using all ancient cities whose locations reflect the economic suitability of their surrounding areas to a greater extent.

We examine ancient port locations, rather than ancient cities, because even those ancient city locations may reflect nearby areas' characteristics and an already-developing urban hierarchy. Ancient ports reflect the highly-local geographic availability of natural harbors in the ancient era, where many more ports were needed than in the modern era and were developed where feasible based on coastal shape. Indeed, the main database of ancient port locations includes some places that are deemed to have been good locations for ports even if there is more limited evidence of their ancient use (and our estimates are not sensitive to including or excluding these places). The ancient port structures are long gone, and even the natural harbors are gone in a fifth of cases, and we verify that our estimates are robust to controlling for proximity to larger ancient cities or ancient roads.

Our empirical strategy draws on the literature on "pointwise" path dependence, which highlights that city locations can be influenced by geographic features that had historical economic relevance. Bleakley and Lin (2012) estimate that "portage sites," where waterway transportation required carrying goods around rapids, became places with persistently higher population density in the United States.[5] Similarly, places with ancient roads are associated with higher modern economic activity (Ahmad and Chicoine, 2021; Franco, Galiani, and Lavado, 2021; Dalgaard et al., 2022; de Benedictis, Licio, and Pinna, 2023), consistent with the "first-stage" emergence of cities at ancient port locations. Fujita and Mori (1996) discuss in theory how cities can form in port locations and continue to thrive after ports themselves become less central to their economic activities, which underlies our "first-stage" empirical result. We then use this development of cities at ancient port locations to explore how these sites influenced the spatial distribution of population in surrounding areas and identify agglomeration shadows.

An advantage of analyzing ancient port locations, which were "seeds" for later cities, is that the local geographic feature that gave rise to an ancient port is plausibly not directly

[5]Bleakley and Lin (2012) also show the emergence of "sister cities" at river mouths downriver from their corresponding portage sites, though we control for distance to the coast and river mouths that have direct geographic advantages.

associated with modern outcomes in places 10-50km away. Even for the ancient port locations themselves, the large majority of these locations have no commercial shipping presence today, and many have no apparent port structures in satellite images. Even when the natural harbor survived, maritime activities are not as central to the modern economy. Ducruet et al. (2024) find zero impact on local population from even large-scale commercial ports in the modern era, though containerization increased population around some ports relative to other ports (Brooks, Gendron-Carrier, and Rua, 2021). Port structures in surrounding locations may compete with each other but can also be complementary; in our context, we do not find evidence of local spatial crowd-out in modern port construction among our ancient port locations.

A related literature explores how economic activity is affected, positively and negatively, by decreased transportation costs within a network (Redding and Sturm, 2008; Faber, 2014; Donaldson and Hornbeck, 2016; Bakker et al., 2021; Flückiger et al., 2021; Barsanetti, 2023; Hornbeck and Rotemberg, 2024). In contrast, we focus on how cities themselves affect surrounding economic activity given these spatial linkages. We focus on the emergence of cities in the local surrounding geography, leaving aside further interacting influences on endogenous city formation across further locations that are intractable (Henderson and Thisse, 2024).

II Data Construction

II.A Ancient Port Locations

We start with a database of ancient port locations, assembled by de Graauw (2019), which has precise geographic coordinates of ports mentioned in ancient texts and histories. The full database lists 4,561 ancient ports across Western and Southern Europe, North Africa, and West Asia. We exclude ports more than 50km from the current coastline and those in remote areas with less complete coverage, which leaves 4,263 ancient ports (Figure 4).

We do not observe other characteristics for these port locations in the ancient era, except that these locations would have had natural harbors.[6] These ancient ports are defined as places that served seafarers rather than just local fishing. Ancient seafarers had a greater need for frequent stops and safe harbors, in contrast to the modern era, when fewer ports are needed.

For each ancient port location, we use Google Earth images to hand-code additional data for the modern period, including whether these locations have a natural harbor. We focus on the presence of natural harbors, setting aside the visible influence of human-made harbor

[6] There are 22 open-water ancient ports, which had more supplementary human-made support and were less reliant on a natural harbor, which we exclude in robustness checks.

protections, and our use of "harbor" refers to "natural harbor."[7] For ancient port locations, 80% retained their natural harbor (Figure 4, Panel B) and 20% lost their natural harbor (Figure 4, Panel C). We also collect data on the presence of modern port structures: 46% have any port structures and 6% have commercial port structures.

The loss of natural harbors was fairly common and widespread, occurring throughout the sample region. Typical reasons for the loss of a natural harbor are long-run natural processes of silting and coastal drift, along with rarer sinking and land rise. Harbor loss occurred even near major ancient cities, such as Rome (Portus, Ostia Antica), Alexandria (Thonis-Heracleion), Leptis Magna, and Caesarea Maritima, as people were unable to resist the natural processes causing harbor loss for long periods even in locations where doing so would have been highly desirable.

II.B Population Data

We use population data for 1km-by-1km grid cells in 2015 from GPWv4 (CIESIN, 2018). These data provide detailed geographic variation in "city" and "urban" activity. We define "city" activity as log population density above 9, or roughly 8,000 people per square km. We define "urban" activity as log population density above 6, or roughly 400 people per square km. We also report impacts on average log population density.[8]

For robustness analysis, we also use other measures of local economic activity. We use grid cell population data for 2000, or GRUMP data (CIESIN, 2011), based on an earlier version of the GPW database that reflects a less developed data model.[9] We also check robustness to using population data from the Global Human Settlement Layer or GHSL (Schiavina, Freire, and MacManus, 2019).[10] Our later simulation exercises also use city locations from GHSL's Urban Centre Database (Florczyk et al., 2019), which provides city centroid coordinates and geographic area along with total population.

Our sample covers 2.3 million grid cells within 50km of the coast and within 200km of their nearest ancient port. As the urban hierarchy was forming in the ancient era, a typical

[7] We use coastal shape to define natural harbors with protected inlets, which is feasible to do by hand for our set of ancient port locations, whereas Gerring et al. (2022) use coastal shape to assign natural harbors for the entire coastline with a polynomial model. Based on satellite images, our measure of natural harbors does not incorporate harbor depth and sea floor topography, but we focus on smaller boats also rather than large-scale commercial uses.

[8] Throughout the paper, "log" refers to the natural logarithm.

[9] The GRUMP dataset uses an older version of GPW data as an input (GPWv3) and then reallocates the population data according to urban-rural distributions within administrative units based on night lights data and settlement population counts. GPWv4 instead uses higher-quality population inputs to construct a grid, with coverage and granularity improvements such that the number of input administrative units increases fivefold. As a result, GPWv4 has nonzero population densities at many more locations.

[10] GHSL re-weights and adjusts the same inputs as GPWv4 with a focus on "built-up" areas and settlements. Because we are interested in the lack of settlement density as well as its presence, which can be sensitive to GHSL decisions about city boundaries, we use the GPWv4 model in our main specifications.

day of travel for pack animals or carts was 12-30km (Scheidel, 2015), and medieval royal itineraries suggest that a day's land travel speed was stable at around 40km per day for many centuries (Hall, 2023). 4,263 grid cells have at least one ancient port, and Appendix Figure A.4 shows the number of grid cells by distance to their nearest ancient port.

II.C Geographic Characteristics and Other Data

Our empirical analysis allows for spatial variation in geographic determinants of cities and economic activity, in contrast to the motivating theoretical framework that assumes a featureless plain. We look to control for "first nature" location characteristics that influence city formation.

Along with restricting our analysis to areas within 50km of the coast, we measure grid cells' distance to the coast, distance to major rivers, and latitude/longitude. We also include terrain ruggedness, based on the standard deviation of elevations within the grid cell, which influenced the locations of economic activity in the region (Nunn and Puga, 2012; Accetturo, Cascarano, and de Blasio, 2019). Our baseline controls also include grid cells' average temperature and precipitation in both January and July.

For robustness analysis, we use a variety of other geographic characteristics. We measure agricultural suitability from the FAO-GAEZ database for six crops: potatoes, cowpeas, olives, chickpeas, wheat, and barley. We also measure whether a grid cell is in a desert. We define grid cells' country or, given the endogenous formation of country boundaries, group cells into 2-degree squares. We measure whether grid cells are on islands and sometimes exclude cells on islands smaller than 2,500km^2.

We also use some data from the ancient era. These include data on the locations of the most important ancient cities and towns taken from the Barrington Atlas (Talbert, 2000; Hanson, 2016): the 14 most important ancient cities in our region ("Category 1"); the 160 next most important ("Category 2"); and 813 less important cities and towns ("Category 3"). We compare the local impacts of ancient ports to these ancient cities, and sometimes exclude areas within 20km of Category 1 ancient cities. We also measure the distance to the nearest Roman road as a potential control variable, though it is endogenous to ancient development.

Appendix A provides additional details on data sources and our database construction.

III Main Estimating Equations

We start by estimating the local impacts of ancient ports on modern outcomes in grid cell i:

$$Y_i = \beta Port_i + \theta \mathbf{X}_i + \varepsilon_i. \tag{1}$$

Outcomes Y_i include log population density and indicators for log population density exceeding 6 (urban activity) or 9 (city activity). The coefficient β reports differences for grid cells with an ancient port, compared to other grid cells with similar geographic characteristics ($\mathbf{X_i}$). Our baseline controls included in $\mathbf{X_i}$ are: log distance to the coast; log distance to the nearest major river; latitude; longitude; terrain ruggedness; and average temperature and precipitation in both January and July.

Our main specification extends equation (1), estimating impacts by distance to nearest ancient ports:

$$\begin{aligned} Y_i &= \textstyle\sum_{d=0km}^{50km} \beta_d^H NearestPort_i^{Harbor} \in [d, d+1) \\ &+ \textstyle\sum_{d=0km}^{50km} \beta_d^{NH} NearestPort_i^{NoHarbor} \in [d, d+1) + \theta \mathbf{X}_i + \varepsilon_i. \end{aligned} \tag{2}$$

The coefficients β_d^H report impacts by distance to the nearest ancient port with a surviving natural harbor, and the coefficients β_d^{NH} report impacts by distance to the nearest ancient port with no surviving natural harbor. Of main interest is how these coefficients compare to each other, whereas their absolute value is estimated relative to the omitted distance categories that are >50km for each port type.

More distant ports could also matter, in principle, but the motivating theoretical framework has clearest predictions for distances from a particular fixed starting location. We jointly estimate these distance bin coefficients for each type of ancient port, because proximity to an ancient port without a harbor is correlated with proximity to an ancient port with a harbor. We include all baseline controls from equation (1).

We graph these non-parametric estimates, by distance bin. We also report the fitted relationship with log distance to nearest ancient port, along with estimated deviations between the non-parametric estimates and this fitted relationship. We focus on distance to these original ancient port locations, even as cities themselves may form and re-form endogenously.

Our baseline specifications report standard errors clustered by 8km-by-8km groups, which reflects typical large-city boundaries in our sample region.[11] By comparison, 4km is the median city radius for cities in our sample region (from GHSL 2019). We allow for spatial correlation across grid cells because cities take up space, whereby one large city can generate measured "city activity" across nearby cells, which could then cause correlation in the error terms across nearby cells from one city's formation. We consider treatment effects at further distances, reflecting agglomeration shadows, but the needed adjustment for inference is about

[11] We cluster by 1/12-degree-by-1/12-degree groups, which are roughly 8km-by-8km squares in the middle of our sample region.

correlated errors within a city. We also report estimates that exclude regions around major ancient cities, and our statistical inference is robust to different adjustments for correlated outcomes across nearby cells.[12]

IV Population Growth and Agglomeration Shadows around Ancient Ports

IV.A Main Estimates

Table 1 reports higher modern population density in grid cells with ancient ports, compared to otherwise geographically-similar grid cells, from estimating equation (1). Grid cells with an ancient port are 1.7 percentage points more likely to have city population density (Panel A, Column 3), which is large relative to the sample mean probability of 0.2 percentage points. Ancient port grid cells are also 12 percentage points more likely to have urban population density (Column 2), compared to a sample mean probability of 7.1 percentage points. Average population density is also 60% higher in ancient port grid cells (Column 1). The impacts of ancient ports are more subtle than the impacts of ancient cities themselves (Panel B), which are potentially more related to the suitability of surrounding areas for city formation, but there is a sufficiently large number of ancient ports for their effects to be precisely estimated.

Figure 5 shows impacts on average population density by distance to ancient ports, from estimating equation (2). The impacts on average population density decline monotonically in distance to ancient ports and remain positive at 50km, relative to further grid cells. The nearby impacts are larger for ancient ports that retained their natural harbor (blue circles), but still substantial for ancient ports that have lost their natural harbor (red circles).

Figure 6 shows impacts on urban density and city density, by distance to ancient ports, which are consistent with the simulated appearance of agglomeration shadows in Section I.

There is a greater likelihood of city density at the ancient port, or within a few kilometers, when the ancient port location has retained its natural harbor (Figure 6, Panel B, blue circles). This probability of city density declines quickly in distance, however, and is lowest around 20km from the ancient port before increasing 0.54 percentage points out to a distance of around 40km. This increase in city probability is smaller than the direct effect in ancient port locations, but large relative to the 0.2 percent average probability of city formation.

When the ancient port location has lost its natural harbor, there is more spatial noise in that starting city location out to 20km before the probability of city density declines (Figure 6, Panel B, red circles). Panel A reports corresponding estimates for urban density, which

[12] We also report standard errors two-way clustered by offset 8km-by-8km groups, shifted by 4km North-South and East-West to allow for spatial correlation across the 8km-by-8km groups, or clustered by 25km-by-25km groups. We also report standard errors adjusted for arbitrary spatial correlation within 4km or 8km (Conley, 1999; Colella et al., 2019).

show more monotonic and continued declines in urban activity by distance to ancient ports.

Panels C and D of Figure 6 report the differences in these estimates ($\beta_d^H - \beta_d^{NH}$), which correspond to the simulated differences in Panels C and D of Figure 3. In our data, as in our simulations based on FKM, there is a distinctive relative dip in the likelihood of city density at intermediate distances (Panel D) and a more flattened relative decline in urban activity (Panel C).

Figure 7 adds a fitted log relationship to the estimates from Figure 6, Panels A and B. The decline in urban activity is roughly logarithmic in distance to the ancient ports (Panel A), whereas the dip in city activity (Panel B) at intermediate distances is more distinctive from a smooth logarithmic decay in distance. The decline in city density is more rapid from 0-20km from the ancient port location, and then this pattern reverses from 20-40km. Panels C and D of Figure 7 report the deviation in the estimates from the fitted log relationship, by distance. When the ancient port more precisely pins down a starting city in that location (i.e., when that ancient port has retained its natural harbor), there is a distinctive decline in the likelihood of city activity at intermediate distances. Once at a further distance, however, there is again a greater likelihood of a city emerging relative to the fitted log relationship.

Table 2 reports a few numbers from Figures 6 and 7, Panel D. From Panel D of Figure 6, the differences in city probability are -0.42 percentage points at 20km and 0.12 percentage points at 40km, as compared to the sample mean probability of 0.20 percentage points (Table 2, Panel A, Columns 1 and 2). This difference of 0.54 percentage points is statistically significant (Column 3) and substantial in magnitude relative to the sample mean of 0.2 percentage points. From Panel D of Figure 7, the deviation from log fit is -0.25 percentage points at 20km and 0.16 percentage points at 40km (Table 2, Panel B).

Cities "seeded" in ancient port locations generate agglomeration shadows, discouraging the formation of cities at intermediate distances. The length of this agglomeration shadow is consistent with a typical day's travel of 12-30km for pack animals or carts in the ancient era (Scheidel, 2015) when the urban hierarchy was forming. Cities were more likely to emerge further away, where there was less direct competition with city activity at the ancient port location.

Section I suggests a heterogeneity exercise, in which agglomeration shadows are more detectable in denser and more homogeneous regions than where regional population is lower and economic activity is more heterogeneous. Appendix Figure A.5 shows agglomeration shadows when restricting our sample to areas within the Roman Empire.[13] By contrast, ag-

[13] We use the maps of the Roman Empire for 117 CE produced by the Digital Atlas of Roman and Medieval Cultures (DARMC), by McCormick and Polk (2017), which reflect the Roman Empire's maximal geographic extent excluding brief territorial holdings at the time in Armenia, Assyria, and Mesopotamia, and later in Caledonia.

glomeration shadows are not as visible outside the Roman Empire (Appendix Figure A.6), which contains 44% of our main sample grid cells but only 26% of the modern population. These areas outside the Roman Empire had lower overall economic activity (at least historically) and fewer cities, which makes agglomeration shadows less detectable. The Roman Empire was also associated with more homogeneous transportation costs and other economic parameters that lessen the influence of wave interference.

Our analysis focuses on the appearance of agglomeration shadows in city formation within a 50km-radius catchment zone. In principle, similar underlying economic forces could influence highly-tradable economic activities at much further distances – and in ways that interact across multiple starting cities – but modeling and estimating such interactions across further distances is daunting. There are also no industry-level data at a fine geographic resolution across the broad region.

IV.B Competition Across Cities versus Across Ports

Cities seeded in ancient port locations cast agglomeration shadows that discourage the formation of other cities nearby. One potential reason for this is that ports themselves make nearby ports less needed and thereby discourage city formation at nearby ports. We find little evidence for port competition, however, and more evidence for city competition, in which cities discourage nearby city formation largely separate from impacts through port structures themselves.

To explore these mechanisms, we restrict our sample to grid cells with ancient ports and estimate how the loss of natural harbors influences that grid cell and surrounding grid cells:

$$Y_i = \beta_1 NoHarbor_i + \beta_2 SurroundingShareNoHarbor_i + \theta \mathbf{X}_i + \varepsilon_i. \tag{3}$$

Outcomes Y_i include whether that grid cell has modern port structures, along with log population density and indicators for log population density exceeding 6 (urban activity) or 9 (city activity). The coefficient β_1 reports differences for ancient port grid cells that have since lost their natural harbor, compared to ancient port grid cells that still have a natural harbor. The coefficient β_2 reports differences when a greater share of ancient port grid cells within 5-50km have lost their natural harbor. We control for the number of other ancient port grid cells within 5-50km, in $\mathbf{X_i}$, along with our other baseline controls (log distance to the coast; log distance to the nearest major river; latitude; longitude; terrain ruggedness; and average temperature and precipitation in both January and July).[14]

Table 3 reports that losing a natural harbor substantially decreases the probability that a grid cell has a modern port structure (Column 1, row 1). However, when surrounding

[14] For this specification, we omit 71 ancient port grid cells that have no other ancient port within 5-50km.

ancient ports have lost their natural harbor, and are thereby less likely to have modern port structures, there is no increase in the likelihood of that grid cell having port structures (Column 1, row 2). This suggests that nearby ports do not systematically crowd-out (or crowd-in) port activity through direct competition (or collaboration).

Table 3 also reports that losing a natural harbor decreases that grid cell's average population density, likelihood of urban activity, and likelihood of city activity (Columns 2-4, row 1). These outcomes are greater, however, when surrounding ancient ports have lost their natural harbor and thereby have lower population density (Columns 2-4, row 2).[15] This suggests that cities themselves are in competition, with increased economic activity in a grid cell due to decreased surrounding economic activity.

IV.C Robustness

The appearance of agglomeration shadows, around cities seeded at ancient port locations, is not sensitive to adjusting for some other sources of spatial variation in economic activity. We explore several types of adjustments to our baseline analysis: controlling for additional grid cell characteristics; sample restrictions; alternative measures of city locations; and adjusted inference for spatial correlation.

A concern would be if grid cells within 10-30km of ancient ports happen to be particularly less suitable for city formation than grid cells closer to (or further from) ancient ports. Our baseline specification controls for log distance to the coast; log distance to the nearest major river; latitude; longitude; terrain ruggedness; and average temperature and precipitation in both January and July. Appendix Figure A.7 includes additional controls for geographic characteristics of the 1km-by-1km grid cells: six crop-specific measures of agricultural suitability (wheat, barley, chickpeas, cowpeas, olives, potatoes);[16] suitability of nearby waters for fishing;[17] log distance to the mouth of a river; indicators for being within 2km, 5km, and 10km of the coast, a river, and the mouth of a river; elevation; and indicators for being in deserts or on islands. Appendix Figure A.8 excludes all grid cells on islands smaller than 2,500 square kilometers (5% of all grid cells). Appendix Figure A.9 reports similar estimates when excluding 216 suitable coastal locations that de Graauw is less sure were used as ancient ports.

We also report estimates that adjust for other features of the ancient economy. Appendix

[15]By contrast, if spatially correlated historical shocks induced local harbor loss and lower local economic activity, this would generate a negative relationship between grid cell economic activity and surrounding harbor loss.

[16]From FAO GAEZ's historical agro-climatic crop suitability data, assuming low inputs (FAO, 2012). The first five crops were important in the ancient era, among those with available crop-specific data, whereas potatoes were influential later (Nunn and Qian, 2011).

[17]Following Dalgaard, Knudsen, and Selaya (2020), we use the average suitability of waters within 100km for 15 common fish species, with data from Aquamaps (2019).

Figure A.10 includes controls for grid cells' log distance to the nearest Roman road, along with log distance to 14 important ancient cities (Barrington 1 sites), although these already reflect responses to the early formation of a geographic urban hierarchy. Appendix Figure A.11 excludes grid cells within 20km of 22 ancient open-water ports, which reflect known human-made investment in artificial harbor breakwaters, and Appendix Figure A.12 excludes grid cells within 20km of the Barrington 1 cities.

Our sample covers a broad geographic area, but our estimates are not sensitive to adjusting for more regional variation in economic activity. Appendix Figure A.13 includes controls for 2-degree by 2-degree fixed effects in cells' latitude and longitude, and Appendix Figure A.14 includes country fixed effects. Country boundaries are themselves endogenous, forming along with the urban hierarchy, so we omit country fixed effects from our baseline specifications.

Our main analysis classifies "city activity" and "urban activity" as grid cells with log population density greater than 9 and 6, respectively. Appendix Figures A.15 and A.16 show the emergence of agglomeration shadows as the population density threshold increases from 5 to 10. We utilize population data from 2015, using version 4 of the GPW model, but estimates are similar using population data from 2000 from an earlier version of the model (Appendix Figure A.17). Estimates are also similar in Appendix Figure A.18 using population data from 2015 from the GHSL-POP model (Schiavina, Freire, and MacManus, 2019). Appendix Figure A.19 reports estimates using city locations from GHSL (Florczyk et al., 2019), where the outcome is being within the radius of a city of population over 500k.

We allow for spatial correlation across grid cells, as the median-size city in our sample region has a radius of 4km and so one shock to city formation can be jointly reflected in nearby cells within the city's boundary. Our baseline inference clusters in 1/12-degree by 1/12-degree groups, or roughly 8km by 8km squares in the middle of our sample region. Appendix Figure A.20 reports similar standard errors when two-way clustering with offset groups, simultaneously clustering by the original groups and new groups shifted by 1/24 of a degree, which relaxes the assumption that grid cells are independent across the original group boundaries. It is computationally intensive to allow for arbitrary spatial correlation around each of the 2.3 million grid cells, following Conley (1999) and Colella et al. (2019), but the statistical inference is similar with arbitrary spatial correlation up to a distance cutoff of 4km (Appendix Figure A.22) or 8km (Appendix Figure A.23). These adjustments for spatial correlation are about the error term itself, due to a city appearing directly in adjacent grid cells, which is separate from potential treatment effects on city formation at further distances. The statistical presence of agglomeration shadows is also not sensitive to increasing the group size to 1/4-degree by 1/4-degree (Appendix Figure A.21), however, or approximately 25km by 25km.

V Modern City Spacing: Realized versus Random

This section considers the general spacing between all modern cities in our sample region, which helps to clarify the null hypothesis and to generalize beyond our earlier analysis around ancient port locations. Our previous analysis explores how the evolved spatial structure responds to particular stimuli over a long time horizon, from the need for many shelter points along the coast in the ancient era, whereas this last exercise characterizes spatial patterns in that evolved urban hierarchy. In this analysis of the modern cross section, we take city locations as given and characterize their spacing, once their size and location have co-evolved with their surroundings.

We now consider the distance between each city and its nearest city and explore whether the realized distribution of minimum distances is distinct from a simulated random distribution. Agglomeration shadows would result in fewer than expected cities with close neighbors, with more than expected nearest neighbors after emerging from that shadow. In focusing on the distribution of minimum city distances, with a distribution-free test based on a simulated random null, this exercise differs from other work in economic geography that characterizes the equilibrium spatial distribution of cities (Dobkins and Ioannides, 2001; Ioannides and Overman, 2004; Rauch, 2014; Mori, Smith, and Hsu, 2020; Henderson, Peng, and Venables, 2022; Mori et al., 2023). Our analysis is more related to a "dartboard" approach to measuring industry agglomeration (Ellison and Glaeser, 1997, 1999; Duranton and Overman, 2005), which compares observed patterns in industry shares to what could be expected randomly, but we consider city locations.

Our analysis here is closest to an approach used in biological statistics to characterize, for example, the spatial distribution of tree seedlings (Diggle, 1983). We adjust this approach for an economic geography context, however. Rather than drawing purely random locations, we predict city likelihood using grid cell characteristics and use the fitted value for each grid cell as a probability weight when drawing different random locations for cities.[18] In particular, this allows for the (random) locations of cities to be systematically close to the coast. More suitable geographic locations for cities, as measured by observable fundamentals, are often positively spatially correlated. If the remaining unobserved characteristics are also (conditionally) positively spatially correlated, the random benchmark would be too spaced

[18] For this prediction, we use a probit model with country fixed effects and the baseline geography controls from equations (1) and (2). This ensures positive weights that are less than one, and restricts city locations to countries with cities in our sample region. Most locations are unlikely to have a city (in an overall sense), but some locations are much more likely to have cities than others (in a relative sense). Without this adjustment, influences on city spacing would be confounded with spatial variation in first-nature geographic advantages.

apart and bias against our detecting agglomeration shadows.

We use GHSL data for urban center locations in our sample region, and implement our analysis separately for cities in three population size categories: over 500k, over 250k, and over 100k. This exercise relies on administrative city boundaries from GHSL, rather than grid cell data, because two nearby dense grid cells could be part of the same city. Cities are not uniformly dense, within their boundaries, but higher population density is correlated with higher city population: the correlation between log population and log population density is 0.37, and cities with greater population are also more likely to have grid cells with log population density above 9.

Figure 8, Panel A, shows that modern large cities are less likely to have neighbors within 40km, and more likely to have (further) neighbors within 40-60km, consistent with agglomeration shadows around large cities and the relative success of cities just beyond those shadows. The red line shows the realized cumulative density function (CDF) of minimum distances between the 66 modern large cities and their nearest large city. The black lines show the CDF from a weighted random draw of 66 city locations in 1,000 simulations, graphing the median along with the 5th centile and 95th centile. The red line is to the right of the thin black line when the actual distribution of modern cities has fewer cities with nearest neighbors within a given distance than 95% of randomly drawn city locations. When the red line slopes up more steeply than the black lines, there are more observed nearest neighbors at that distance than the random simulations.

Panels C and E of Figure 8 show that agglomeration shadows are less pronounced when including smaller cities. The observed CDF is somewhat distinct from random simulations at some initial distances, before increasing more rapidly, but this difference is less substantial than for the largest cities.

Using defined "city" locations to measure agglomeration shadows raises a potential measurement concern: the area just outside a defined city may not be another city because, if it were, it may have been included in that first defined city.[19] We can adjust for this effect by shifting the simulated distribution of cities, however, using the variation in city sizes to construct non-overlapping circles around cities: if we randomly draw another city within that circle, we drop it and draw another city.[20] This shifts the simulated distributions to the

[19]That is, cities may indeed develop near each other but become considered one large city: when the locations of city boundaries and city centers are defined ex post, along with the concurrent development of surrounding areas, these definitions could overstate agglomeration shadows. For example, if the boundary of a large city is drawn such that it contains nearby densely populated areas, then areas just outside that boundary will mechanically have low population density.

[20]For each city, we take the square root of its area over π to define the radius of a circle that approximates that city's geographic footprint. When redrawing cities, we randomly assign each real radius to each drawn city. Drawing is done sequentially, such that we draw a city randomly, then randomly assign a radius, and

right, as shown in Panels B, D, and F of Figure 8. There are still fewer-than-expected large cities within 40km of other large cities and more catch-up between 40-60km (Panel B), but there are no longer detectable agglomeration shadows for cities with population above 250k or 100k (Panels D and F).

These results are consistent with agglomeration shadows around large cities, despite the potential for spatially-correlated location fundamentals and particular shocks accumulated over the course of millennia that might lead to more close-by cities than we would expect by chance.

VI Conclusion

The location of economic activity has direct implications for landowners and, since people are generally not perfectly mobile, also for individuals' well-being more generally. Indeed, a motivation for place-based policies is to help people in targeted locations by encouraging local growth. Locally successful policy may discourage growth nearby, however, when concentrations of economic activity compete with and discourage rival centers caught in their "agglomeration shadow."

We focus on identifying agglomeration shadows using ancient ports as seeds from which cities emerged. The locations of ancient ports, which were shaped by very local coastline features, provide empirical traction to explore impacts on nearby places. Ancient port locations may themselves continue to benefit from their local geography, but these locations are plausibly uncorrelated with a down-and-up wave in unobservable grid cell characteristics between 20km and 40km away. Further, where natural harbors survived in these ancient port locations, the location of modern cities was more precisely pinned down, mitigating the problem of wave interference. Where harbors did not survive, even with strong agglomeration shadows, the alternating peaks and valleys in economic activity can average out and appear as smooth declines in economic activity.

We estimate agglomeration shadows in city formation, which appear as a distinctive non-monotonic down-and-up wave in city formation. Large cities are less likely to form at intermediate distances from ancient ports, caught in the agglomeration shadow of cities seeded at ancient port locations. Cities are likelier to form further away, however, just beyond those shadows. This contrasts with a more general monotonic decline in population density and urban activity in distance to the ancient port, which is consistent with nearby places becoming confined to second-tier status within an urban hierarchy. Loss of nearby harbors does not increase locations' own port activity in the modern era, but it does encourage local

then redraw if, for any other already-drawn city, the distance between the new city and the already-drawn city is less than the sum of their radii. Otherwise, this drawn city is kept and its radius cannot be assigned again.

economic activity by decreasing nearby cities.

We show that the appearance of agglomeration shadows is not mechanical, but that the spacing between all large cities is statistically distinct from random. There are fewer large cities at intermediate distances than we would expect at random, after accounting for locational fundamentals. In characterizing the cross-sectional spatial distribution of economic activity, this analysis complements our study of the emergence of city spacing in response to particular stimuli from ancient port locations.

Early subsidized investments in local economic activity can generate sustained long-run local growth, when there are positive local agglomeration spillovers, and we extend this literature to identify further impacts on the spatial organization of economic activity in the surrounding urban hierarchy. When economic activity clusters in a location, agglomeration forces also generate an "agglomeration shadow" that discourages economic activity in nearby locations until sufficient distance makes rival centers viable.

References

Accetturo, Antonio, Michele Cascarano, and Guido de Blasio. 2019. "Pirate Attacks and the Shape of the Italian Urban System." Working Paper.

Ahlfeldt, Gabriel M., and Nicolai Wendland. 2013. "How Polycentric is a Monocentric City? Centers, Spillovers and Hysteresis." *Journal of Economic Geography*, 13(1): 53–83.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica*, 83(6): 2127–2189.

Ahmad, Zofia, and Luke Chicoine. 2021. "Silk Roads to Riches: Persistence Along an Ancient Trade Network." Working Paper.

Ali, Kamar, M. Rose Olfert, Mark D. Partridge, and Dan S. Rickman. 2009. "Do New Economic Geography Agglomeration Shadows Underlie Current Population Dynamics across the Urban Hierarchy?" *Papers in Regional Science*, 88(2): 445–467.

Allen, Treb, and Dave Donaldson. 2022. "Persistence and Path Dependence in the Spatial Economy." Working Paper.

Alonso, William. 1964. *Location and Land Use: Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

Bakker, Jan David, Stephan Maurer, Jörn-Steffen Pischke, and Ferdinand Rauch. 2021. "Of Mice and Merchants: Connectedness and the Location of Economic Activity in the Iron Age." *Review of Economics and Statistics*, 103(4): 652–665.

Barsanetti, Bruno. 2023. "Road Endpoints and City Sizes." *Review of Economics and Statistics*, 1–45.

Beltrán Tapia, Francisco J., Alfonso Díez-Minguela, and Julio Martinez-Galarraga. 2021. "The Shadow of Cities: Size, Location and the Spatial Distribution of Population." *Annals of Regional Science*, 66: 729–753.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." *Quarterly Journal of Economics*, 127(2): 587–644.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. 2019. "What Drives Differences in Management Practices?" *American Economic Review*, 109(5): 1648–1683.

Bosker, Maarten, and Eltjo Buringh. 2017. "City Seeds: Geography and the Origins of the European City System." *Journal of Urban Economics*, 98: 139–157.

Brooks, Leah, Nicolas Gendron-Carrier, and Gisela Rua. 2021. "The Local Impact of Containerization." *Journal of Urban Economics*, 126.

Center for International Earth Science Information Network (CIESIN). 2011. "Global Rural-Urban Mapping Project (GRUMPv1): Population Density Grid." Palisades, NY: NASA Socioeconomic Data and Applications Center. Version 1. Accessed 2021.

Center for International Earth Science Information Network (CIESIN). 2018. "Gridded Population of the World (GPWv4): Population Density." Palisades, NY: NASA Socioeconomic Data and Applications Center. Version 4, Revision 11. Accessed 2021.

Christaller, Walter. 1933. *Die Zentralen Orte in Süddeutschland.* Translated in 1966 by C.W. Baskin to The Central Places in Southern Germany. Englewood Cliffs, NJ: Prentice-Hall.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan, and Mathias Thoenig. 2019. "Inference with Arbitrary Clustering." *IZA Working Paper*, 1(12584).

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1): 1–45.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams. 2019. "The More We Die, the More We Sell? A Simple Test of the Home-Market Effect." *Quarterly Journal of Economics*, 134(2): 843–894.

Cuberes, David, Klaus Desmet, and Jordan Rappaport. 2021. "Urban Growth Shadows." *Journal of Urban Economics*, 123: 103334.

Cummins, Steven, and Sally Macintyre. 2002. "'Food Deserts' — Evidence and Assumption in Health Policy Making." *BMJ*, 325(7361): 436–438.

Dalgaard, Carl-Johan, Anne Sofie B. Knudsen, and Pablo Selaya. 2020. "The Bounty of the Sea and Long-run Development." *Journal of Economic Growth*, 25(3): 259–295.

Dalgaard, Carl-Johan, Nicolai Kaarsen, Ola Olsson, and Pablo Selaya. 2022. "Roman Roads to Prosperity: Persistence and Non-Persistence of Public Infrastructure." *Journal of Comparative Economics*, 50(4): 896–916.

Davis, Donald R., and Jonathan I. Dingel. 2020. "The Comparative Advantage of Cities." *Journal of International Economics*, 123: 103291.

de Benedictis, Luca, Vania Licio, and Anna Maria Pinna. 2023. "From the Historical Roman Road Network to Modern Infrastructure in Italy." *Journal of Regional Science*, 63(5): 1162–1191.

de Graauw, Arthur. 2019. "The Catalogue of Ancient Coastal Settlements, Ports and Harbours." Accessed 2019. https://www.ancientportsantiques.com/.

Diggle, Peter J. 1983. *Statistical Anaylsis of Spatial Point Patterns.* London: Academic Press.

Dingel, Jonathan I., Joshua D. Gottlieb, Maya Lozinski, and Pauline Mourot. 2023. "Market Size and Trade in Medical Services." Working Paper.

Dobkins, Linda Harris, and Yannis M. Ioannides. 2001. "Spatial Interactions Among US Cities: 1900–1990." *Regional science and Urban Economics*, 31(6): 701–731.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *Quarterly Journal of Economics*, 131(2): 799–858.

Ducruet, César, Réka Juhász, Dávid Krisztián Nagy, and Claudia Steinwender. 2024. "All Aboard: The Effects of Port Development." Conditionally Accepted.

Duranton, Gilles, and Diego Puga. 2004. “Micro-Foundations of Urban Agglomeration Economies.” In *Handbook of Regional and Urban Economics*. Vol. 4, 2063–2117. Amsterdam: Elsevier.

Duranton, Gilles, and Diego Puga. 2023. “Urban Growth and Its Aggregate Implications.” *Econometrica*, 91(6): 2219–2259.

Duranton, Gilles, and Henry G. Overman. 2005. “Testing for Localization Using Micro-Geographic Data.” *Review of Economic Studies*, 72(4): 1077–1106.

Ellison, Glenn, and Edward L. Glaeser. 1997. “Geographic Concentration in US Manufacturing Industries: A Dartboard Approach.” *Journal of Political Economy*, 105(5): 889–927.

Ellison, Glenn, and Edward L. Glaeser. 1999. “The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?” *American Economic Review*, 89(2): 311–316.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns.” *American Economic Review*, 100(3): 1195–1213.

Faber, Benjamin. 2014. “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System.” *Review of Economic Studies*, 81(3 (288)): 1046–1070.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. “Place-Based Drivers of Mortality: Evidence from Migration.” *American Economic Review*, 111(8): 2697–2735.

Florczyk, A., C. Corbane, M. Schiavina, M. Pesaresi, L. Maffenini, M. Melchiorri, P. Politis, F. Sabo, S. Freire, D. Ehrlich, T. Kemper, P. Tommasi, D. Airaghi, and L. Zanchetta. 2019. “GHS-UCDB R2019A: GHS Urban Centre Database 2015, Multitemporal and Multidimensional Attributes.” Brussels: European Commission, Joint Research Centre (JRC), Global Human Settlement Layer (GHSL). Accessed 2021.

Flückiger, Matthias, Erik Hornung, Mario Larch, Markus Ludwig, and Allard Mees. 2021. “Roman Transport Network Connectivity and Economic Integration.” *Review of Economic Studies*, 89(2): 774–810.

Franco, Ana Paula, Sebastian Galiani, and Pablo Lavado. 2021. “Long-Term Effects of the Inca Road.” Working Paper.

Fujita, Masahisa, and Paul R. Krugman. 1995. “When is the Economy Monocentric?: Von Thünen and Chamberlin Unified.” *Regional Science and Urban Economics*, 25(4): 505–528.

Fujita, Masahisa, and Tomoya Mori. 1996. “The Role of Ports in the Making of Major Cities: Self-Agglomeration and Hub-Effect.” *Journal of Development Economics*, 49(1): 93–120.

Fujita, Masahisa, and Tomoya Mori. 2005. “Frontiers of the New Economic Geography.” *Papers in Regional Science*, 84(3): 377–405.

Fujita, Masahisa, Paul R. Krugman, and Anthony Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge, MA: MIT press.

Fujita, Masahisa, Paul R. Krugman, and Tomoya Mori. 1999. "On the Evolution of Hierarchical Urban Systems." *European Economic Review*, 43(2): 209–251.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114(3): 739–767.

Gerring, John, Brendan Apfeld, Tore Wig, and Andreas Forø Tollefsen. 2022. *The Deep Roots of Modern Democracy: Geography and the Diffusion of Political Institutions.* Cambridge, UK: Cambridge University Press.

Giroud, Xavier, Simone Lenzu, Quinn Maingi, and Holger Mueller. 2021. "Propagation and Amplification of Local Productivity Spillovers." Working Paper.

Glaeser, Edward L., and Joshua D. Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economics and Spatial Equilibrium in the United States." *Journal of Economic Literature*, 47(December): 983–1028.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118(3): 536–598.

Hall, Jacob. 2023. "Travel Speed over the Longue Durée." Working Paper.

Hanson, J.W. 2016. "Cities Database." Oxford: Oxford Roman Economy Project (OXREP). Version 1.0. Accessed July 2020. http://www.romaneconomy.ox.ac.uk/.

Helpman, Elhanan, and Paul Krugman. 1985. *Market Structure and Foreign Trade.* Cambridge, MA: MIT Press.

Henderson, J. Vernon. 1974. "The Sizes and Types of Cities." *American Economic Review*, 64(4): 640–656.

Henderson, J. Vernon, and Jacques-Francois Thisse. 2024. "Urban and spatial economics after 50 years." *Journal of Urban Economics*, 144: 1–13.

Henderson, J. Vernon, Cong Peng, and Anthony Venables. 2022. "Growth in the African Urban Hierarchy." Working Paper.

Hornbeck, Richard, and Martin Rotemberg. 2024. "Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies." *Journal of Political Economy*, Forthcoming.

International Institute for Applied Systems Analysis (IIASA) and Food and Agriculture Organization of the United Nations (FAO). 2020. "Global Agro-Ecological Zones (GAEZ)." Laxenburg, Austria: IIASA; Rome: FAO. Version 4.0. Accessed June 2024.

Ioannides, Yannis M., and Henry G. Overman. 2004. "Spatial Evolution of the US Urban System." *Journal of Economic Geography*, 4(2): 131–156.

Jefferson, Mark. 1939. "The Law of the Primate City." *Geographical Review*, 29(2): 226–232.

Kaschner, K., K. Kesner-Reyes, C. Garilao, J. Segschneider, T. Rius-Barile, J. Rees, and R. Froese. 2019. "AquaMaps: Predicted Range Maps for Aquatic Species." Accessed May 2024.

Krugman, Paul R. 1995. *Development, Geography, and Economic Theory.* Cambridge, MA: MIT Press Books.

Lösch, August. 1940. *Die Räumliche Ordnung der Wirtschaft.* Translated in 1954 by Jena Fischer to The Economics of Location. New Haven, CT: Yale University Press.

Marshall, Alfred. 1890. *Principles of Economics.* London: Macmillan and Co.

Matsuyama, Kiminori. 2017. "Geographical Advantage: Home Market Effect in a Multi-Region World." *Research in Economics*, 71(4): 740–758.

McCormick, Michael, and Matthew Polk. 2017. "Boundaries: Provinces, ca. 117 CE." Cambridge, MA: The Digital Atlas of Roman and Medieval Civilizations. Version 2023. Accessed August 2024. https://darmc.harvard.edu/.

Michaels, Guy, and Ferdinand Rauch. 2018. "Resetting the Urban Network: 117–2012." *Economic Journal*, 128(608): 378–412.

Mills, Edwin S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review*, 57(2): 197–210.

Mori, Tomoya, and Tony E Smith. 2015. "On the Spatial Scale of Industrial Agglomerations." *Journal of Urban Economics*, 89: 1–20.

Mori, Tomoya, Takashi Akamatsu, Yuki Takayama, and Minoru Osawa. 2023. "Origin of Power Laws and Their Spatial Fractal Structure for City-Size Distributions." Working Paper.

Mori, Tomoya, Tony E. Smith, and Wen-Tai Hsu. 2020. "Common Power Laws for Cities and Spatial Fractal Structures." *Proceedings of the National Academy of Sciences*, 117(12): 6469–6475.

Muth, Richard. 1969. *Cities and Housing : The Spatial Pattern of Urban Residential Land Use.* Chicago: University of Chicago Press.

Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics*, 94(1): 20–36. Accessed replication data September 2019. https://diegopuga.org/data/rugged/.

Nunn, Nathan, and Nancy Qian. 2011. "The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment." *Quarterly Journal of Economics*, 126(2): 593–650.

Plato. 360 B.C.E.. *Plato in Twelve Volumes.* Vol. 1, Translated by Harold North Fowler in 1966. Cambridge, MA: Harvard University Press.

Rauch, Ferdinand. 2014. "Cities as Spatial Clusters." *Journal of Economic Geography*, 14(4): 759–773.

Redding, Stephen J, and Daniel M. Sturm. 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review*, 98(5): 1766–1797.

Scheidel, Walter. 2015. "Orbis: The Stanford Geospatial Network Model of the Roman World." Working Paper.

Schiavina, Marcello, Sergio Freire, and Kytt MacManus. 2019. "GHS-POP R2019A - GHS Population Grid Multitemporal (1975-2015)." Brussels: European Commission, Joint Research Centre (JRC), Global Human Settlement Layer (GHSL). Accessed 2021.

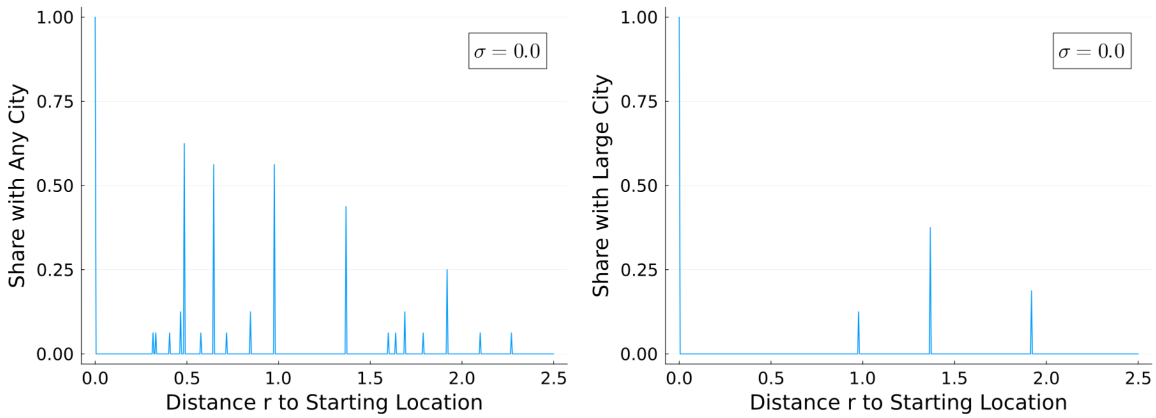
Talbert, Richard L.A., ed. 2000. *Barrington Atlas of the Greek and Roman World.* Princeton, NJ: Princeton University Press.

Tervo, Hannu. 2010. "Cities, Hinterlands and Agglomeration Shadows: Spatial Developments in Finland During 1880–2004." *Explorations in Economic History*, 47(4): 476–486.

von Thünen, Johann Heinrich. 1826. *Der Isolierte Staat.* Edited by Peter Hall. Translated in 1966 by Carla M. Wartenberg to Von Thünen's Isolated State: An English Edition of Der Isolierte Staat. Oxford: Pergamon Press.

Figure 1. Share of Instances with Cities by Distance to Fixed Starting City

Panel A. Any City (≥1 Industry) Panel B. Large City (≥2 Industries)

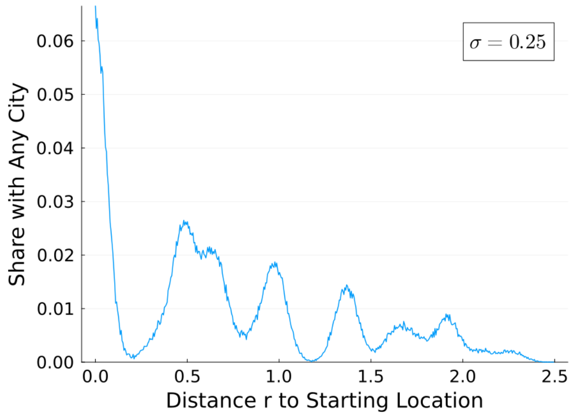


Notes: This figure shows the share of instances in which there is a city at the indicated distance r from a fixed starting city, as modeled for the 16 aggregate population parameter values in Figure 7 of Fujita, Krugman, and Mori (1999). Four of these 16 values are bifurcation values, where cities are "born" or "die" or change their number of industries, and we use city locations on the right side of the bifurcation associated with larger population.

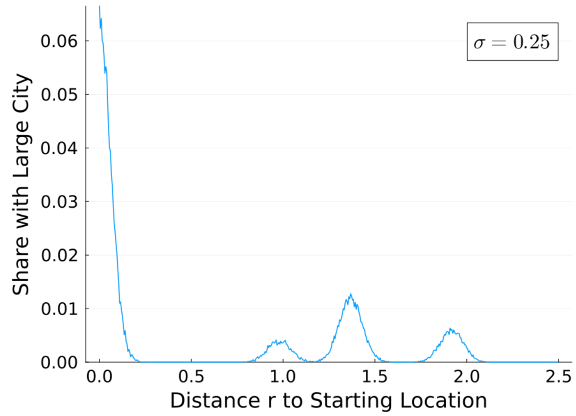
Panel A shows the share of instances with any city at that distance r, defined as a location with at least one manufacturing industry. Panel B shows the share of instances with a large city at that distance r, defined as a location with at least 2 manufacturing industries. The distance r is restricted to the positive direction, along this one-dimensional economy, from the location of the fixed initial city at distance 0.

Figure 2. Share of Simulations with Cities by Distance to Fixed Starting Location, with Increasing Spatial Noise

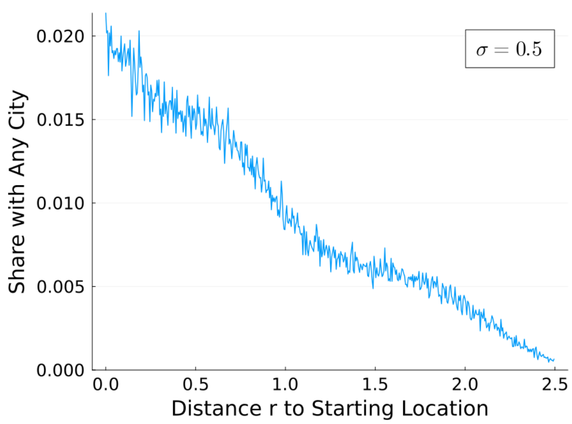
Panel A. Any City, with Low Noise



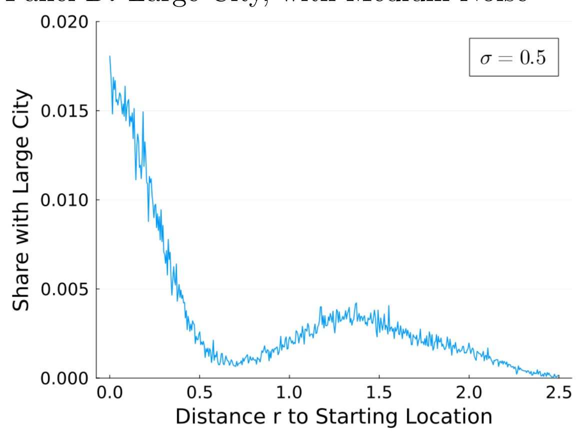
Panel B. Large City, with Low Noise



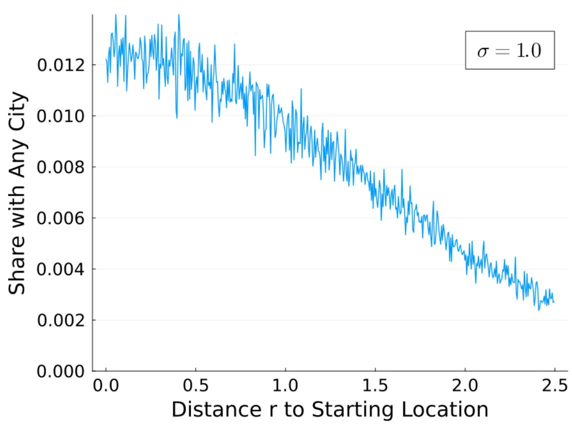
Panel C. Any City, with Medium Noise



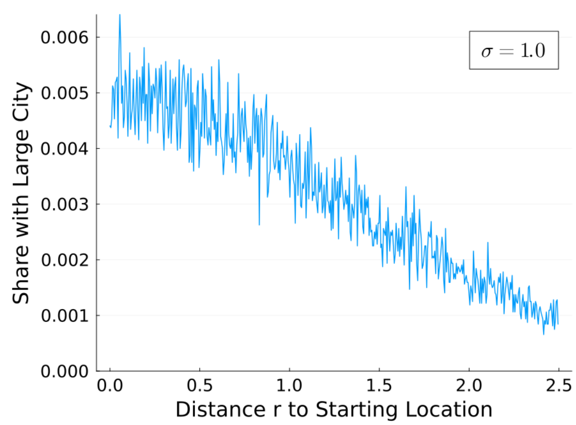
Panel D. Large City, with Medium Noise



Panel E. Any City, with High Noise



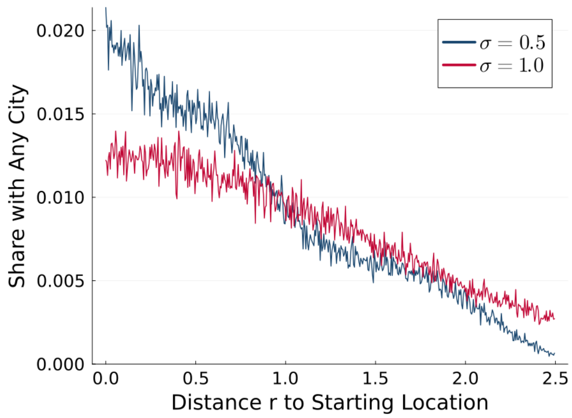
Panel F. Large City, with High Noise



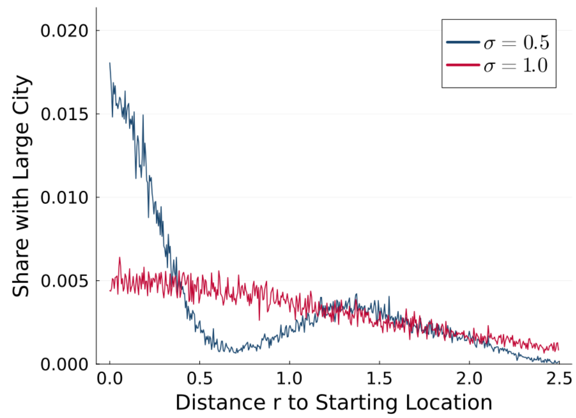
Notes: Each panel reports the share of simulations in which there is a city at the indicated distance r from a fixed starting location. In each of 2,000 simulations, we shift the city locations from Figure 1 left or right by a spatial noise term drawn from a mean zero normal distribution with increasing standard deviation: "Low Noise" ($\sigma = 0.25$) in Panels A and B; "Medium Noise" ($\sigma = 0.5$) in Panels C and D; and "High Noise" ($\sigma = 1.0$) in Panels E and F. The distance r is restricted to the positive direction from the location of the fixed starting location at distance 0 (which can now differ from the location of that previously central initial city).

Figure 3. Share of Simulations with Cities by Distance: Medium versus High Noise

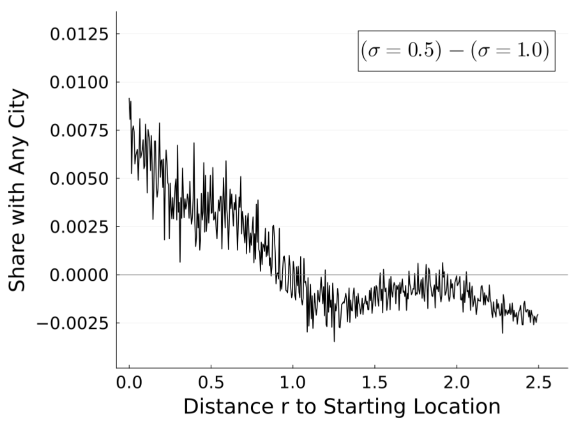
Panel A. Any City, Medium and High Noise



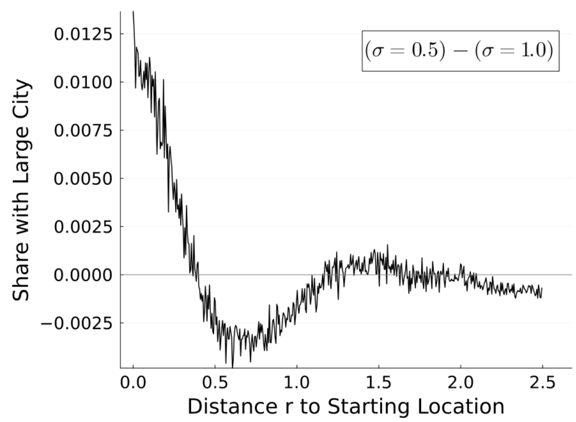
Panel B. Large City, Medium and High Noise



Panel C. Any City, Difference between Medium and High Noise



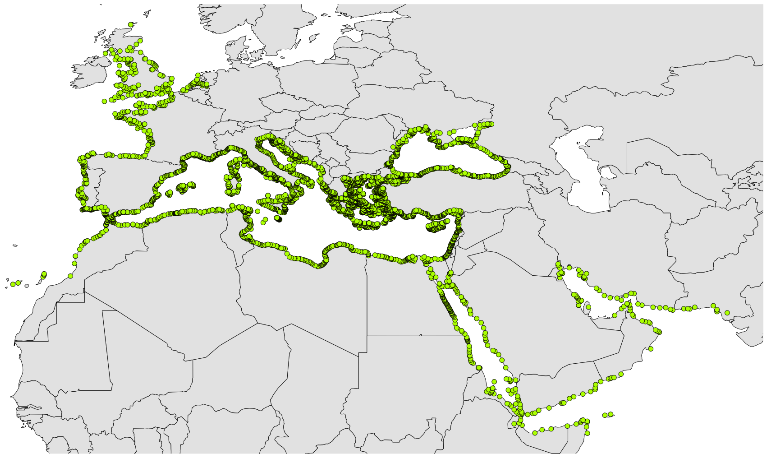
Panel D. Large City, Difference between Medium and High Noise



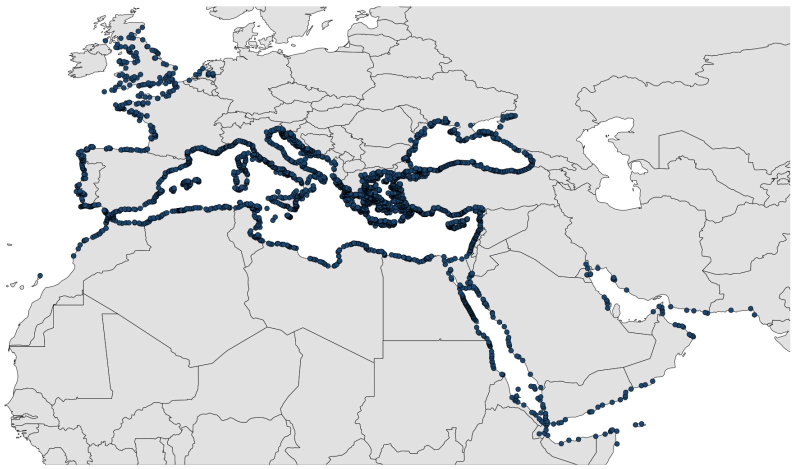
Notes: This figure compares the "Medium Noise" and "High Noise" cases from Figure 2. Panels A and B show simulations with Medium and High Noise on the same plot, and Panels C and D show the difference in shares (Medium Noise minus High Noise).

Figure 4. Ancient Port Locations in Our Main Sample

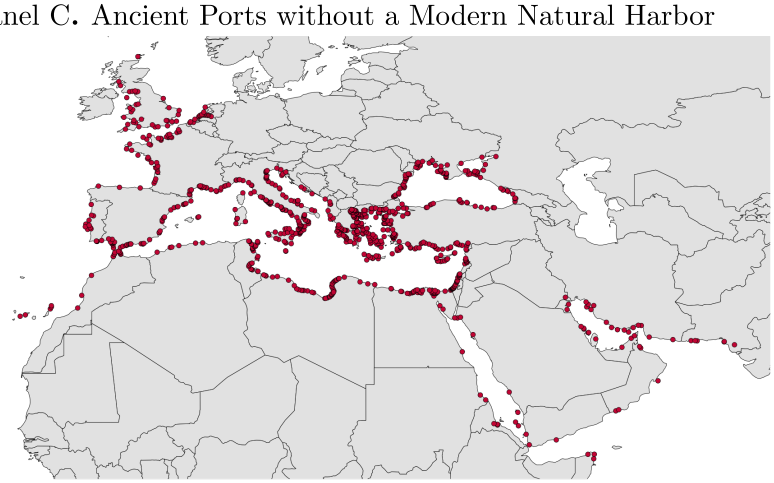
Panel A. All Ancient Ports



Panel B. Ancient Ports with a Modern Natural Harbor

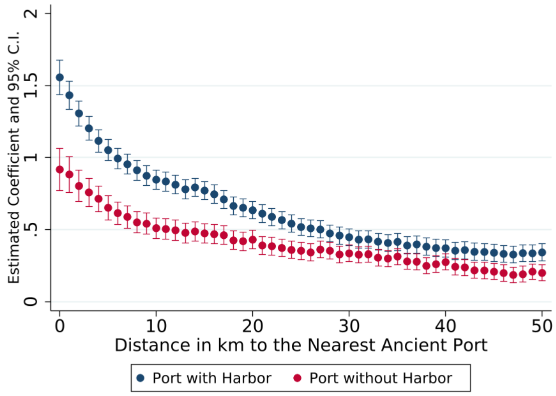


Panel C. Ancient Ports without a Modern Natural Harbor



Notes: This figure shows the locations of ancient ports in our main sample. Panel A shows all ancient ports, Panel B shows ancient port locations that have a natural harbor in modern satellite images, and Panel C shows ancient port locations that do not have a natural harbor in modern satellite images.

Figure 5. Impacts on Average Log Population Density, by Distance to Nearest Ancient Port

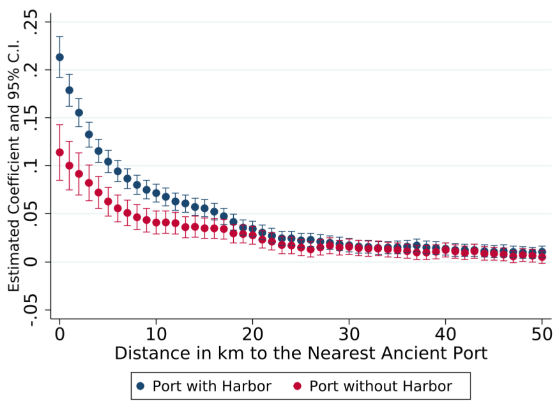


Notes: This figure shows impacts on average log density by distance to the nearest ancient port with and without a natural harbor, from estimating equation (2). The sample is 1km-by-1km grid cells within 50km of the coast, and include the baseline controls: log distance to the coast, log distance to the nearest major river, latitude, longitude, terrain ruggedness, and average temperature and precipitation in both January and July.

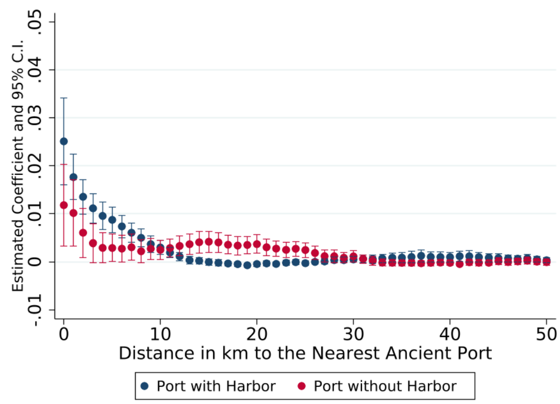
Vertical lines report 95% confidence intervals, based on robust standard errors clustered at 1/12-degree-by-1/12-degree groups (roughly 8km-by-8km in the middle of our sample region).

Figure 6. Impacts on Probability of Urban Density or City Density, by Distance to Nearest Ancient Port

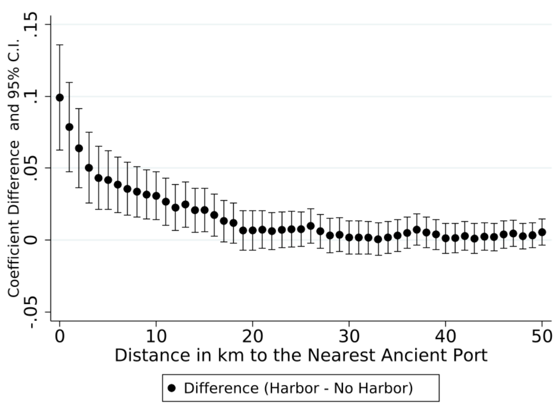
Panel A. Probability of Urban Density



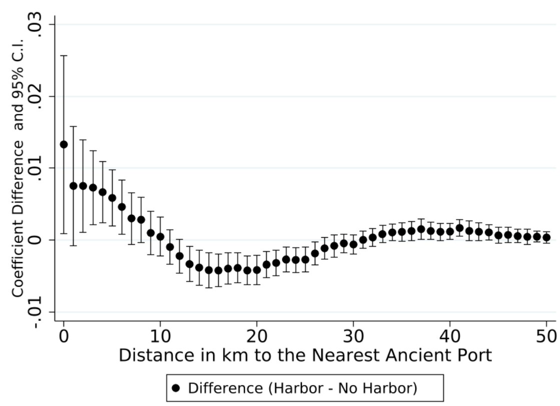
Panel B. Probability of City Density



Panel C. Differences for Urban Density



Panel D. Differences for City Density



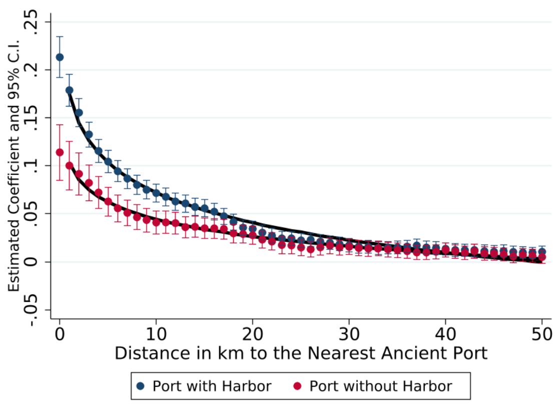
Notes: This figure shows impacts on the probability of a grid cell having "Urban Density" (ln population density > 6) or "City Density" (ln population density > 9), by distance to the nearest ancient port with and without a natural harbor, from estimating equation (2). The sample is 1km-by-1km grid cells within 50km of the coast, and include the baseline controls: log distance to the coast, log distance to the nearest major river, latitude, longitude, terrain ruggedness, and average temperature and precipitation in both January and July.

Panels A and B show coefficients on distance to the nearest ancient port with a modern natural harbor (blue circle) and without a modern natural harbor (red circle). Panels C and D show the differences between these estimates (with harbor minus without harbor).

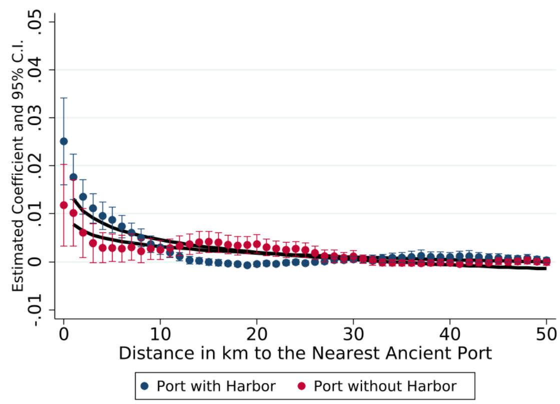
Vertical lines report 95% confidence intervals, based on robust standard errors clustered at 1/12-degree-by-1/12-degree groups (roughly 8km-by-8km in the middle of our sample region).

Figure 7. Impacts on Probability of Urban Density or City Density, with Fitted Log Relationship

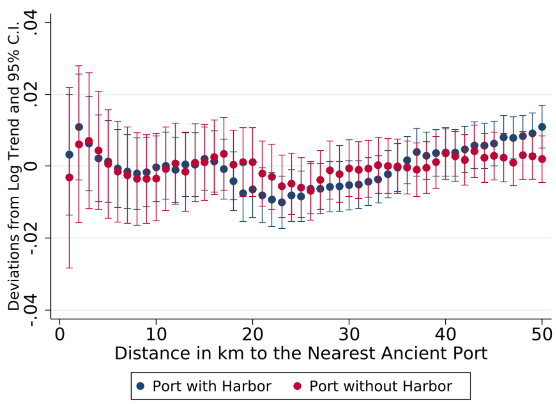
Panel A. Probability of Urban Density



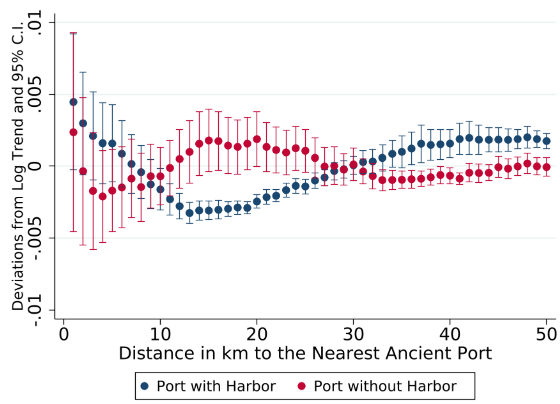
Panel B. Probability of City Density



Panel C. Urban, Difference from Log Fit



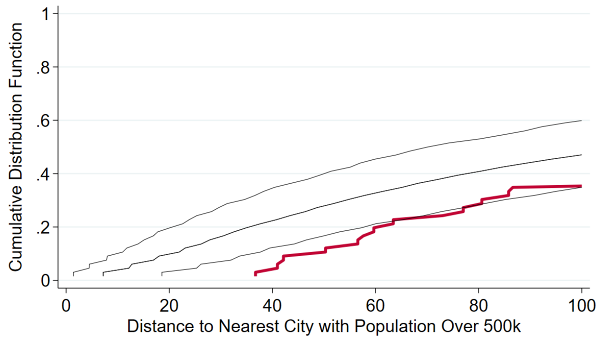
Panel D. City, Difference from Log Fit



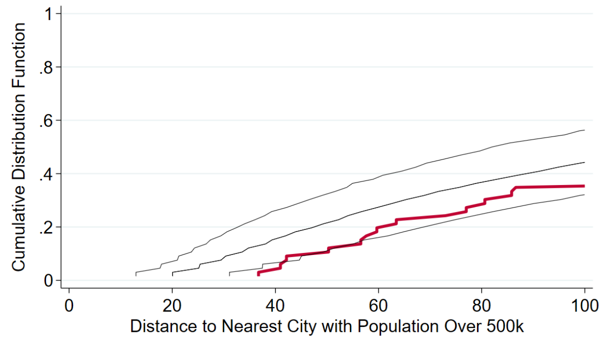
Notes: Panels A and B reproduce Panels A and B from Figure 6, with the addition of two black lines that represent the fitted log relationships from regressing the coefficients from distance bins 1–50 on log distance. Panels C and D report the coefficients from Panels A and B, subtracting the fitted log relationships (black lines). Vertical lines report 95% confidence intervals using the standard errors from Panels A and B.

Figure 8. Modern City Spacing: Realized versus Random

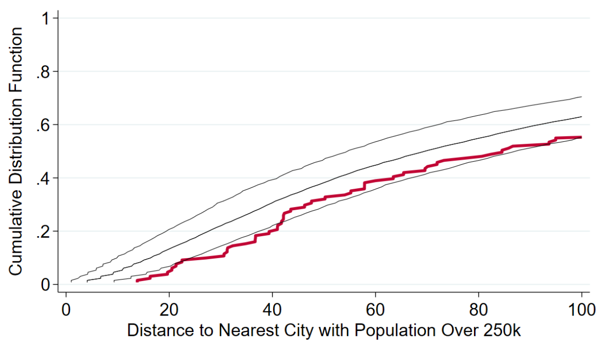
Panel A. Cities with Population above 500k



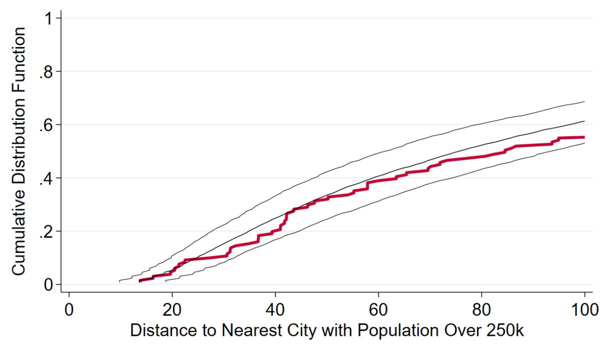
Panel B. Cities with Population above 500k, with Area Adjustment



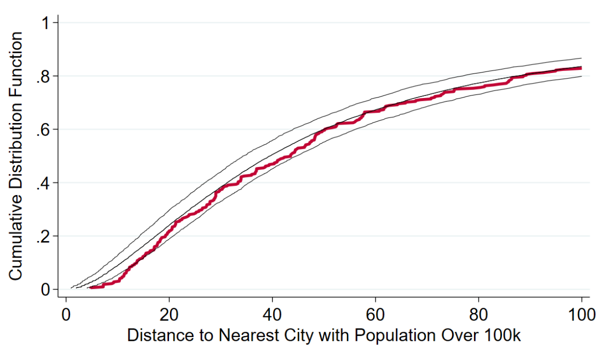
Panel C. Cities with Population above 250k



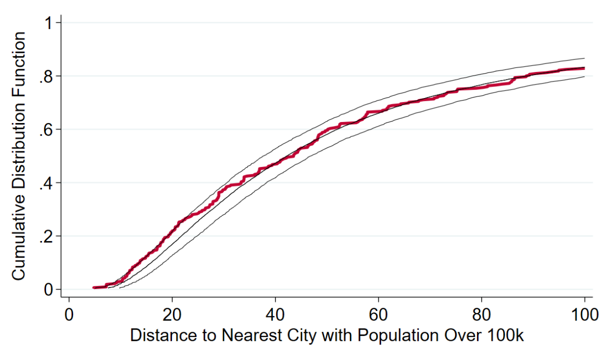
Panel D. Cities with Population above 250k, with Area Adjustment



Panel E. Cities with Population above 100k



Panel F. Cities with Population above 100k, with Area Adjustment



Notes: This figure compares real and simulated distributions of distances between nearest cities, by city size. The red line shows the real distribution of distances to cities. The central gray line shows the median simulated distance, and the upper and lower gray lines show the 5th and 95th percentiles. We identify real city locations and their 2015 populations using GHSL data. The distributions of random city locations are randomly drawn 1,000 times from our baseline grid sample. For each city size category, the random draws use probability weights for each grid cell based on its fitted value from a probit model predicting city likelihood with country fixed effects and the baseline controls: log distance to the coast, log distance to the nearest major river, latitude, longitude, terrain ruggedness, and average temperature and precipitation in both January and July. For Panels B, D, and F, the "Area Adjustment" to the simulated black lines refers to when cities are redrawn if they are within the geographic footprint of an already-placed city, as described in Section V. The figure is truncated at 100km.

Table 1. Estimated Impacts of Ancient Ports on Grid Cell Population Density

	Ln Population Density (1)	Urban Density (2)	City Density (3)
Panel A.			
Ancient Port	0.589	0.116	0.017
	(0.044)	(0.008)	(0.004)
Panel B.			
Barrington 1 City	3.867	0.663	0.395
	(0.675)	(0.124)	(0.155)
Barrington 2 City	2.423	0.418	0.114
	(0.217)	(0.048)	(0.032)
Barrington 3 City	1.714	0.308	0.038
	(0.098)	(0.022)	(0.009)
Outcome Mean	3.290	0.071	0.002
	[1.888]	[0.257]	[0.045]
Observations	2,278,861	2,278,861	2,278,861

Notes: This table shows the relationship between ancient features and modern population densities, from estimating equation (1). Column 1 reports estimated impacts on grid cell log population density, Column 2 reports estimated impacts on a grid cell having "Urban Density" (log population density > 6), and Column 3 reports impacts on a grid cell having "City Density" (log population density > 9). We measure population density using 2015 population per square kilometer estimates from the GPWv4 model (CIESIN, 2018).

Panel A shows effects of an ancient port site. For comparison, Panel B shows effects of ancient cities in three size categories, using ancient city data from Talbert (2000) and Hanson (2016). In our sample, restricted to 1km-by-1km grid cells within 50km of the coast, there are 10 cities in the 1st category, 101 in the second category, and 480 in the third category.

All estimates include the baseline controls for grid cell characteristics: log distance to the coast, log distance to the nearest major river, latitude, longitude, terrain ruggedness (standard deviation of elevations within the grid cell), and average temperature and precipitation in both January and July. In parentheses are robust standard errors clustered by 1/12-degree-by-1/12-degree groups (roughly 8km-by-8km in the middle of our sample region). For the outcome means, standard deviations are reported in square brackets.

Table 2. Probability of City Density at Distances to Ancient Port

	At 20km (1)	At 40km (2)	Difference: 40km - 20km (3)
Panel A. With Natural Harbor vs. Without (from Figure 6, Panel D)			
Estimated Difference	-0.0042	0.0012	0.0054
	(0.0010)	(0.0006)	(0.0013)
Panel B. With Natural Harbor vs. Log Fit (from Figure 7, Panel D)			
Estimated Deviation from Log Fit	-0.0025	0.0016	0.0040
	(0.0002)	(0.0005)	(0.0005)
Outcome Mean	0.0020	0.0020	0.0020
	[0.0447]	[0.0447]	[0.0447]
Observations	2,278,861	2,278,861	2,278,861

Notes: This table reports a few numbers from Figures 6 and 7, Panel D. Panel A corresponds to Panel D of Figure 6, and Panel B corresponds to Panel D of Figure 7. Column 3 reports the difference between the coefficients in Columns 2 and 1.

Table 3. Estimated Impacts on Ancient Port Grid Cells from Harbor Loss and Nearby Ancient Port Grid Cells Losing Their Natural Harbor

	Modern Port Structure (1)	Ln Population Density (2)	Urban Density (3)	City Density (4)
No Harbor	-0.383	-0.269	-0.050	-0.017
	(0.016)	(0.099)	(0.019)	(0.008)
Surrounding Share with No Harbor	-0.039	0.973	0.134	0.041
	(0.039)	(0.225)	(0.044)	(0.020)
Outcome Mean	0.467	4.721	0.267	0.024
	[0.499]	[2.162]	[0.443]	[0.154]
Observations	4,198	4,198	4,198	4,198

Notes: This table shows effects on ancient port grid cells from losing their own natural harbor and nearby ancient port grid cells losing their natural harbor. These coefficients are estimated and reported jointly, following equation 3, for the effect of having no modern natural harbor and "Surrounding Share with No Harbor" that refers to the share of ancient port grid cells within 5-50km that have no modern natural harbor. This sample is restricted to grid cells with an ancient port, within 50km of the coast, that also have at least one other ancient port grid cell within 5-50km.

Column 1 reports effects on having a "Modern Port Structure," which is an indicator for satellite-visible human-made structures to assist the loading and unloading of boats (including a basic pier). Column 2 reports effects on log population density of each grid cell, Column 3 reports effects on "Urban Density" (log population density > 6), and Column 4 reports effects on "City Density" (log population density > 9).

All estimates control for the number of other ancient port grid cells within 5-50km and the baseline controls for grid cell characteristics: log distance to the coast, log distance to the nearest major river, latitude, longitude, terrain ruggedness (standard deviation of elevations within the grid cell), and average temperature and precipitation in both January and July. In parentheses are robust standard errors clustered by 1/12-degree-by-1/12-degree groups (roughly 8km-by-8km in the middle of our sample region). For the outcome means, standard deviations are reported in square brackets.

Online Appendix

Identifying Agglomeration Shadows: Long-run Evidence from Ancient Ports

Richard Hornbeck
University of Chicago

Guy Michaels
London School of Economics

Ferdinand Rauch
University of Heidelberg

June 2025

A Data Appendix

A.1 Ancient Ports Data

The database on ancient port locations comes from de Graauw (2019). Our use of "ancient" refers to a period around 1,500-3,500 years ago, largely in the Classical World, and limited to the Mediterranean and surrounding areas. The term "port" refers to artificial structures to assist the loading and unloading of boats. De Graauw geo-located these ports based on ancient and modern sources, including the Barrington Atlas, Pleiades dataset, and the Digital Atlas of the Roman Empire (DARE). A large majority of these ancient ports are Greek and Roman. These ancient ports are locations that were used by seafarers sailing over long distances, rather than locations for local fishermen who may have landed their boats on beaches.

The artificial nature of a port distinguishes it from a harbor (or harbour in British English), which is a place where maritime vessels can seek shelter. Almost all of the ancient ports cataloged by de Graauw (2019) are presumed to have been constructed in locations that had a natural harbor. We received from de Graauw a list of 22 known open-water ports from his database that relied more on human-made protections from the sea, which we sometimes exclude from our analysis. Our baseline analysis includes 216 suitable locations for ancient ports, which de Graauw is less certain of their use as ancient ports, and the estimates are not sensitive to excluding these locations.

Some natural harbors have disappeared due to natural processes over the many centuries since those ancient times. The most common of these natural processes was silting and coastal drift, though in some places sinking and land rising also occurred.

We employed the firm Digital Divide Data (DDD) to extract visual information on the modern locations of the ancient ports. We obtained a satellite image for each ancient port location, using Google Maps, and DDD recorded information visible in these images. The main variables constructed this way are indicators for a natural harbor in the vicinity of each ancient port location (within approximately 1km); modern port structures in that location; and whether these modern port structures (where they exist) are extensive structures, as opposed to basic ones (e.g., a simple pier).

A.2 Other Data Sources

We obtain population data from the NASA Center for International Earth Science Information Network (CIESIN, 2018). We use data for 2015, obtained from Version 4, Revision 11 of the Gridded Population of the World dataset (GPWv4) by CIESIN, with robustness analysis using data for 2000 from CIESIN's earlier Global Rural-Urban Mapping Project, Version 1 (CIESIN, 2011). Both of these datasets provide estimates of population density for a grid of 30 arc seconds, a resolution that is slightly finer than our main 1km-by-1km grid, which we discuss below.

We also use two datasets from the "Global Human Settlement Layer" (GHSL). For data on modern cities, we use the GHS Urban Centre Database, which has coordinates of city centers, total population, and geographic areas in 2015 (Florczyk et al., 2019). These GHSL data capture the location and geographic extent of modern cities, but rely on defined geographic boundaries of cities and so are more dependent on those chosen administrative boundaries as compared to the more-flexible GPWv4 population density data. We also obtain the "GHS population grid" dataset for 2015 population (Schiavina, Freire, and MacManus, 2019), which provides an alternative measure of grid-level population density but also with a focus on defined "built-up" areas.

We use maps from ESRI for the coastline, river locations, and modern country borders (Esri Atlas, 2014; Esri Data and Maps, 2010). We use these maps to assign grid cells to their country and to calculate distance to the coast. An indicator for deserts comes from the NASA Earth Observatory land cover map (NASA Earth Observatory, 2020), and elevation data come from the Global Digital Elevation Model Version 3 (NASA and METI, 2019).

Other data on grid cells' geographic characteristics comes from several sources. We obtain data on temperature and precipitation from the WorldClim 2 Database (Fick and Hijmans, 2017). We use the 10-year averages of temperature and precipitation in January and July. For a small share of cells with missing temperature and precipitation, we compute a distance-weighted measure of the temperature and precipitation of their neighbors using inverse distance weighted (IDW) interpolation. We use data on terrain ruggedness, based on the standard deviation of elevations, from Nunn and Puga (2012). We obtain data on crop suitability from FAO-GAEZ (IIASA/FAO, 2012).[21] We obtain crop suitability data for wheat, barley, olives, cowpeas, white potatoes, and chickpeas. We construct our measure of sea bounty with the Aquamaps data for 15 fish species' suitability (Kaschner et al., 2019), reconstructing the bounty of the sea index in Dalgaard, Knudsen, and Selaya (2020).

We use mapped ancient locations from the Barrington Atlas (Talbert, 2000), using the database of cities compiled by Hanson (2016) based on the Barrington data. These cover the cities of the Roman Empire between 100 B.C.E. and C.E. 300. For the locations of Roman roads, we use data from the Digital Atlas of Roman and Medieval Civilizations, or DARMC (McCormick et al., 2013). DARMC also used the Barrington Atlas (Talbert, 2000) to encode the boundaries of the 117 C.E. provinces of the Roman Empire (McCormick and

[21] We measure crop suitability as the agro-climatic potential yield of each crop in dry weight kilogram per hectare, with historical data for 1961-1990 from CRUTS32, assuming rain-fed water supply and low input levels.

Polk, 2017).

A.3 Grid Construction

We merge the above datasets into an equal area grid of 1km^2 cells using the Lambert Azimuthal Equal Area projection, where we place the reference point in the (approximate) center of the Mediterranean at coordinates 39N and 18.5E (following Bakker et al. (2021)). While these grid cells are all 1km^2 in area, they are only approximately 1km-by-1km squares in two dimensions, and so we measure all distances in the paper using grid cell centroid coordinates and their geodesic distances. By construction, there are no measurable distortions of distance or angle in the vicinity of the reference point. Even at the Western end of the Mediterranean (at Gibraltar), the distance between grid cell centroids deviates less than 2% from 1km.

To construct our sample of ancient ports, we begin with the ancient ports database (de Graauw, 2019), which covers ancient ports around the Mediterranean, the European Atlantic, the Black Sea, the Red Sea, the Gulf of Aden, the Gulf of Oman, the Persian Gulf, and a few nearby coastal areas. Of the 4,561 ancient ports in the de Graauw dataset, we exclude 134 as the cumulative result of three geographic exclusion criteria. First, following exchanges with de Graauw, we exclude a few isolated ancient ports in areas where the coverage is sparse (and most likely incomplete).[22] Second, we exclude ancient ports that are assigned to a broad area, such as "Ireland" or "Corsica", rather than to specific coordinates. Third, since we focus on analyzing coastal areas, we exclude ancient ports that are more than 50km from the coast.[23]

We match each ancient port and Barrington city location to a grid cell, based on its nearest grid cell centroid (up to 1.5km away), to allow for slight imprecision in coastal maps and location coordinates. Due to small differences in coastal boundaries across datasets, for grid cells with an ancient port or Barrington city and missing population density, we use the mean population density of its "king's neighbor" grid cells with non-missing values. Following this adjustment, there are 4,333 ancient ports in grid cells with population density. A small number of grid cells contain multiple ancient ports, such that there are 4,263 grid cells with an ancient port and population density.[24] For Table 1, we use an indicator for that grid cell having any ancient port. For the Figures and Table 2, we classify the grid cell

[22] This omits four isolated ancient ports along the German, Scandinavian, and Baltic coasts to the east of Amsterdam. We also exclude one isolated port in Iceland. Along the West African Coast, we restrict the analysis to locations from the Canary Islands northwards. Along the East African coast, we exclude ports south of the Gulf of Aden. We include the coast of the Indian Ocean going east until we reach India.

[23] The excluded inland ports were mostly along the Nile and its delta and a few other major rivers.

[24] There are 4,204 grid cells in our sample with one ancient port, 51 grid cells with two ports, 6 grid cells with three ports, 1 grid cell with four ports, and one grid cell with 5 ports, bringing the number of port grid cells to 4,263.

as having a natural harbor if any location within that grid cell has a natural harbor. For Table 3, similarly, we use an indicator for that grid cell having any modern port structure.

Across our broader sample region, we calculate the distance from each grid cell to its nearest ancient port, with and without a modern natural harbor, and exclude grid cells that are further than 200km from an ancient port. To calculate the distance from each grid cell to the nearest river, coast, and Roman road, we compute the shortest geodesic distance between features and the grid cell centroid.[25] We exclude grid cells more than 50km from the coast. We also restrict the grid cells to modern countries that have at least one ancient port, which means our grid cells follow the coastlines and end at the borders of Netherlands-Germany, India-Pakistan, and Ireland-UK, as well as at the same dividing lines in the Atlantic Maghreb and Somalia.

[25]For other data, we overlay the input maps with the grid cells database and take the values that each grid cell centroid falls on.

Appendix References

Bakker, Jan David, Stephan Maurer, Jörn-Steffen Pischke, and Ferdinand Rauch. 2021. "Of Mice and Merchants: Connectedness and the Location of Economic Activity in the Iron Age." *Review of Economics and Statistics*, 103(4): 652–665.

Center for International Earth Science Information Network (CIESIN). 2011. "Global Rural-Urban Mapping Project (GRUMPv1): Population Density Grid." Palisades, NY: NASA Socioeconomic Data and Applications Center. Version 1. Accessed 2021.

Center for International Earth Science Information Network (CIESIN). 2018. "Gridded Population of the World (GPWv4): Population Density." Palisades, NY: NASA Socioeconomic Data and Applications Center. Version 4, Revision 11. Accessed 2021.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan, and Mathias Thoenig. 2019. "Inference with Arbitrary Clustering." *IZA Working Paper*, 1(12584).

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1): 1–45.

Dalgaard, Carl-Johan, Anne Sofie B. Knudsen, and Pablo Selaya. 2020. "The Bounty of the Sea and Long-run Development." *Journal of Economic Growth*, 25(3): 259–295.

de Graauw, Arthur. 2019. "The Catalogue of Ancient Coastal Settlements, Ports and Harbours." Accessed 2019. https://www.ancientportsantiques.com/.

Esri Atlas. 2014. "World Boundaries and Places." Redlands, CA: Environmental Systems Research Institute (ESRI), ArcGIS Living Atlas of the World. Accessed April 2018.

Esri Data and Maps. 2010. "World Major Rivers." Redlands, CA: Environmental Systems Research Institute (ESRI) Mapping Center, ArcGIS Living Atlas of the World. Accessed 2019.

Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1km Spatial Resolution Climate Surfaces for Global Land Areas." International Journal of Climatology 37 (12): 4302-4315. Accessed January 2020.

Florczyk, A., C. Corbane, M. Schiavina, M. Pesaresi, L. Maffenini, M. Melchiorri, P. Politis, F. Sabo, S. Freire, D. Ehrlich, T. Kemper, P. Tommasi, D. Airaghi, and L. Zanchetta. 2019. "GHS-UCDB R2019A: GHS Urban Centre Database 2015, Multitemporal and Multidimensional Attributes." Brussels: European Commission, Joint Research Centre (JRC), Global Human Settlement Layer (GHSL). Accessed 2021.

Fujita, Masahisa, Paul R. Krugman, and Tomoya Mori. 1999. "On the Evolution of Hierarchical Urban Systems." *European Economic Review*, 43(2): 209–251.

Hanson, J.W. 2016. "Cities Database." Oxford: Oxford Roman Economy Project (OXREP). Version 1.0. Accessed July 2020. http://www.romaneconomy.ox.ac.uk/.

International Institute for Applied Systems Analysis (IIASA) and Food and Agriculture Organization of the United Nations (FAO). 2020. "Global Agro-Ecological Zones (GAEZ)." Laxenburg, Austria: IIASA; Rome: FAO. Version 4.0. Accessed June 2024.

Kaschner, K., K. Kesner-Reyes, C. Garilao, J. Segschneider, T. Rius-Barile, J. Rees, and R. Froese. 2019. "AquaMaps: Predicted Range Maps for Aquatic Species." Accessed May 2024.

McCormick, Michael, and Matthew Polk. 2017. "Boundaries: Provinces, ca. 117 CE." Cambridge, MA: The Digital Atlas of Roman and Medieval Civilizations. Version 2023. Accessed August 2024. https://darmc.harvard.edu/.

McCormick, Michael, Guoping Huang, Giovanni Zambotti, and Jessica Lavash. 2013. "Roman Road Network." Cambridge, MA: The Digital Atlas of Roman and Medieval Civilizations (DARMC). Version 2008. Accessed September 2019. https://darmc.harvard.edu/.

National Aeronautics and Space Administration (NASA), and Ministry of Economy, Trade, and Industry (METI) of Japan. 2019. "Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 3 (GDEM 003)." NASA Earthdata. Accessed June, 2024. https://asterweb.jpl.nasa.gov/gdem.asp.

National Aeronautics and Space Administration (NASA) Earth Observatory. 2020. "Land Cover Classification Map." NASA Earth Observing System Data and Information System (EOSDIS) Land Processes Distributed Active Archive Center. Accessed July 2020. https://earthobservatory.nasa.gov/.

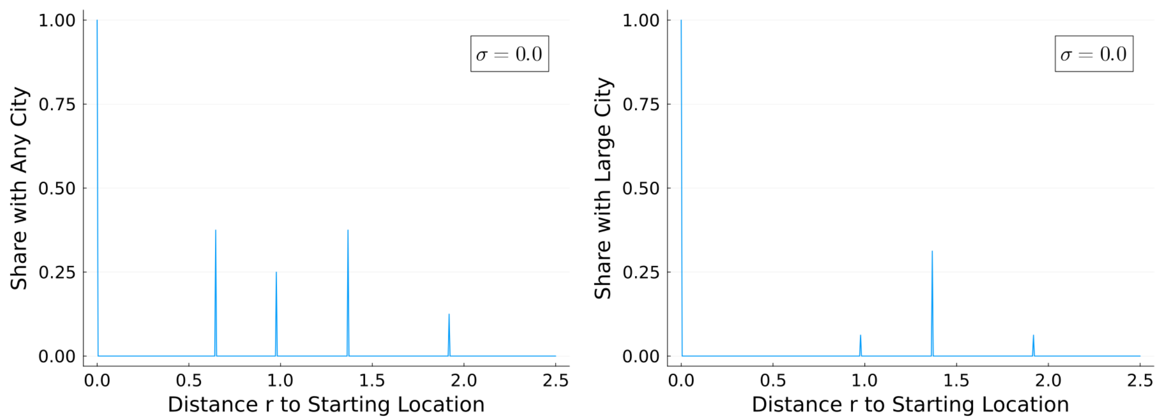
Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics*, 94(1): 20–36. Accessed replication data September 2019. https://diegopuga.org/data/rugged/.

Schiavina, Marcello, Sergio Freire, and Kytt MacManus. 2019. "GHS-POP R2019A - GHS Population Grid Multitemporal (1975-2015)." Brussels: European Commission, Joint Research Centre (JRC), Global Human Settlement Layer (GHSL). Accessed 2021.

Talbert, Richard L.A., ed. 2000. *Barrington Atlas of the Greek and Roman World.* Princeton, NJ: Princeton University Press.

Figure A.1. Share of Instances with Cities by Distance and Population Density

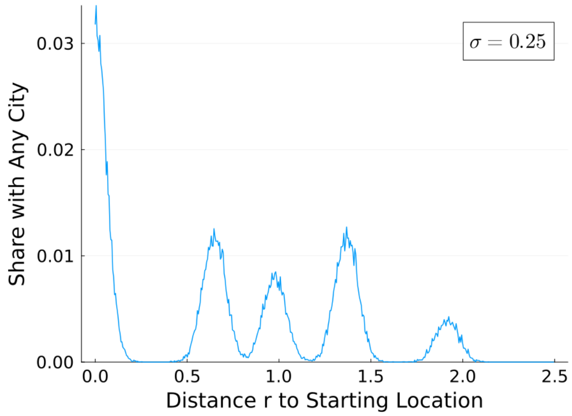
Panel A. Medium Density City, or Greater   Panel B. High Density City



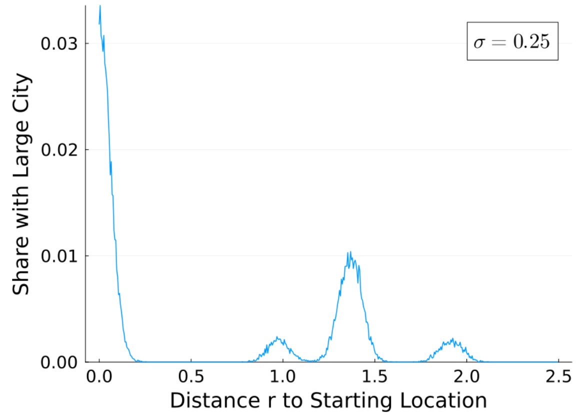
Notes: This figure shows the share of instances in which there is a city at the indicated distance from a starting city, as modeled for the 16 aggregate population parameter values in Fujita, Krugman, and Mori (1999). Panel A shows the share of instances with a city of medium density, defined as a city with log population density > 2, where location is binned to the nearest 0.005. Panel B shows the share of instances with a city of high density, defined as a city with log population density > 3.

Figure A.2. Share of Instances with Cities by Distance and Population Density, with Increasing Spatial Noise

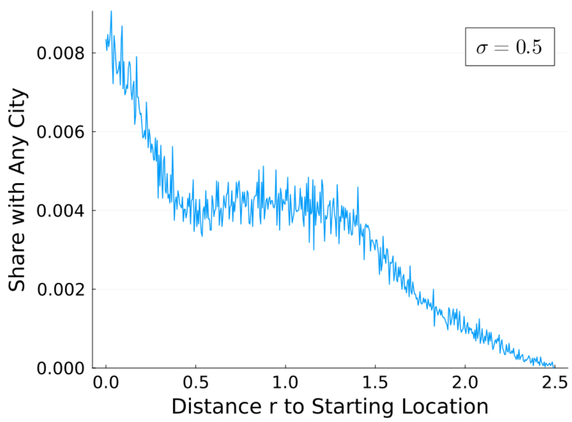
Panel A. Medium Density City, Low Noise



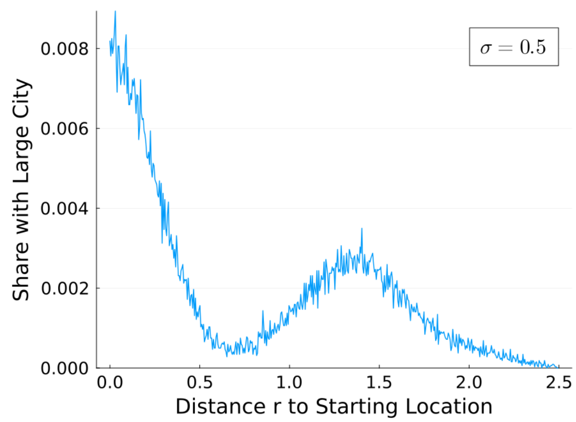
Panel B. High Density City, Low Noise



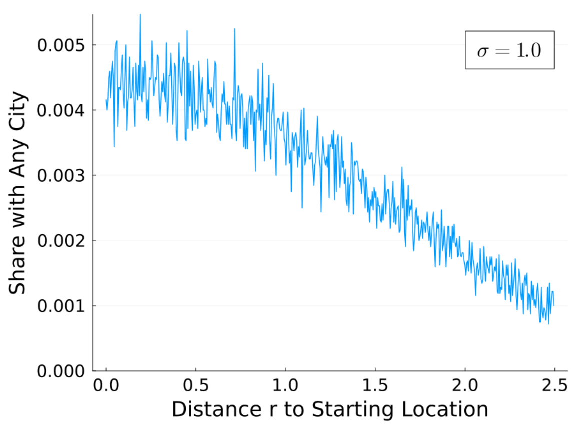
Panel C. Medium Density City, Medium Noise



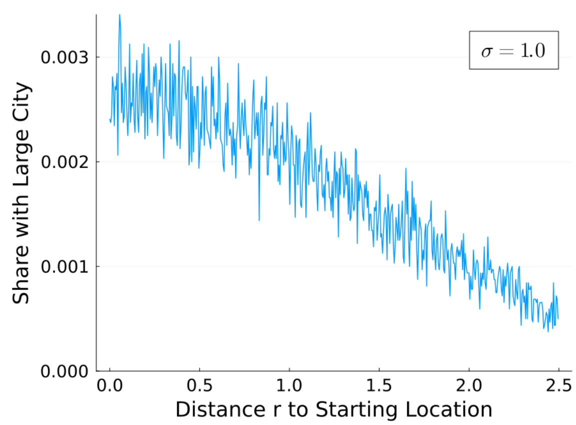
Panel D. High Density City, Medium Noise



Panel E. Medium Density City, High Noise



Panel F. High Density City, High Noise

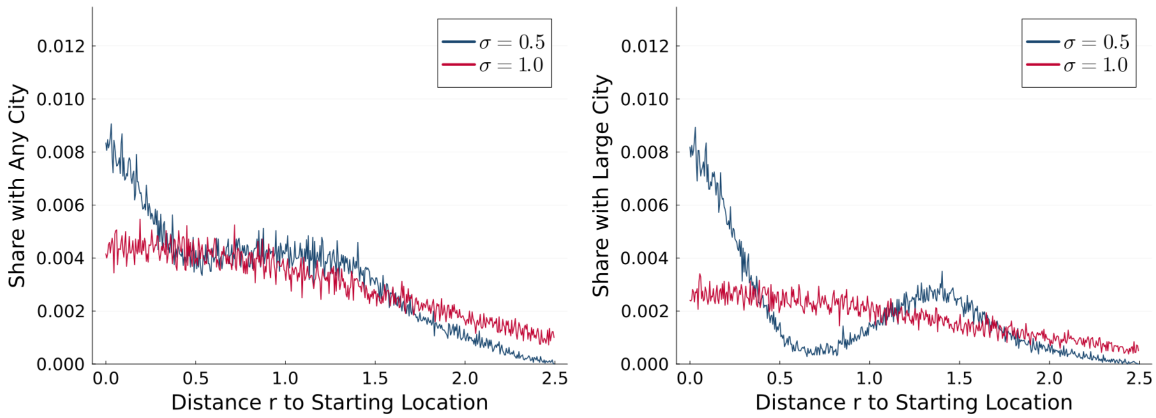


Notes: Each panel reports simulations from Figure A.1, with increasing spatial noise in the fixed starting city's location (at distance zero in Figure A.1): "Low Noise" (standard deviation of 0.25); "Medium Noise" (standard deviation of 0.5); and "High Noise" (standard deviation of 1.0).

Figure A.3. Share of Instances with Cities by Distance and Population Density: Medium versus High Noise

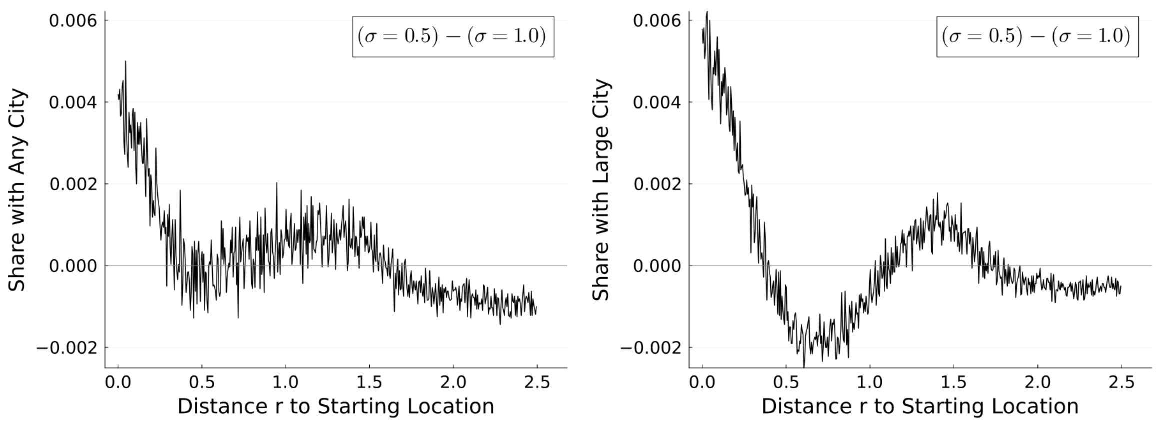
Panel A. Medium Density City, Medium and High Noise

Panel B. High Density City, Medium and High Noise



Panel C. Medium Density City, Difference between Medium and High Noise

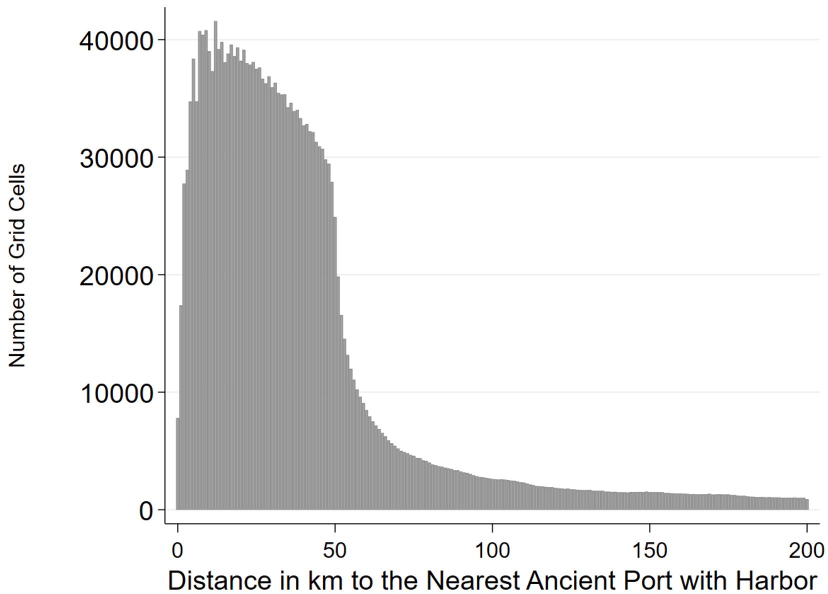
Panel D. High Density City, Difference between Medium and High Noise



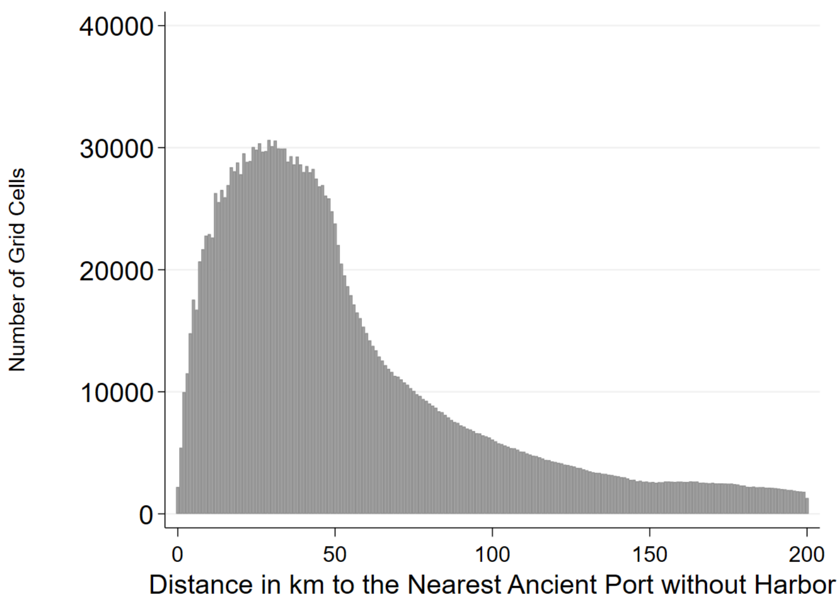
Notes: This figure compares "Medium Noise" and "High Noise" results from the simulations in Figure A.2. Panels A and B show simulations with Medium and High Noise on the same plot, and Panels C and D show the difference in probabilities (Medium Noise minus High Noise).

Figure A.4. Number of Grid Cells by Distance to Nearest Ancient Port, by Modern Harbor Type

Panel A. Distance to Nearest Ancient Port with Harbor



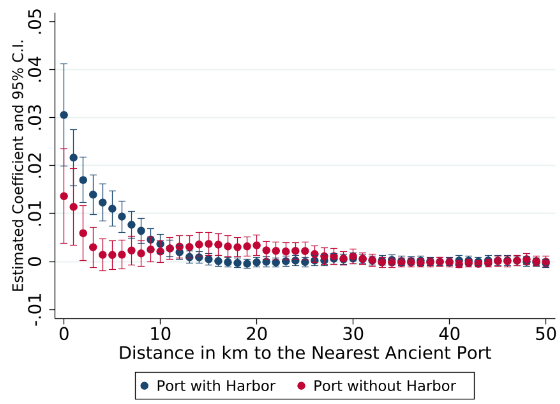
Panel B. Distance to Nearest Ancient Port without Harbor



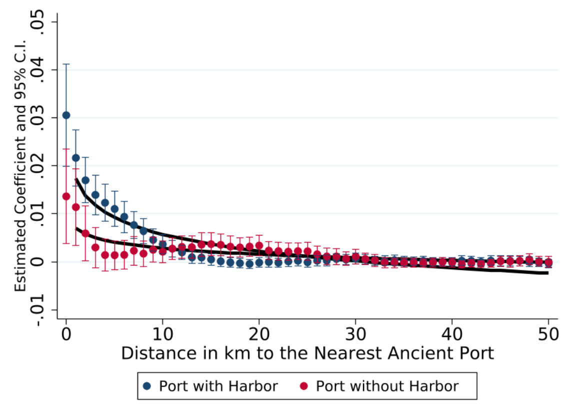
Notes: This figure shows the number of grid cells by distance to nearest ancient port. For each integer kilometer, Panel A shows the number of grid cells at that nearest distance from an ancient port with a natural harbor in the modern era, and Panel B shows the number of grid cells at that distance from a port without a natural harbor in the modern era. We classify a location as having a natural harbor in the modern era if at least one ancient port location at that grid cell is classified as having a natural harbor using satellite imagery. Our sample is restricted to grid cells within 200km of an ancient port, and truncated from this figure are: 2.8% of grid cells that are more than 200km from an ancient port with a modern harbor, and 5.4% of grid cells are more than 200km from an ancient port without a modern harbor.

Figure A.5. Heterogeneous Impacts on Probability of City Density: Only Within the Roman Empire

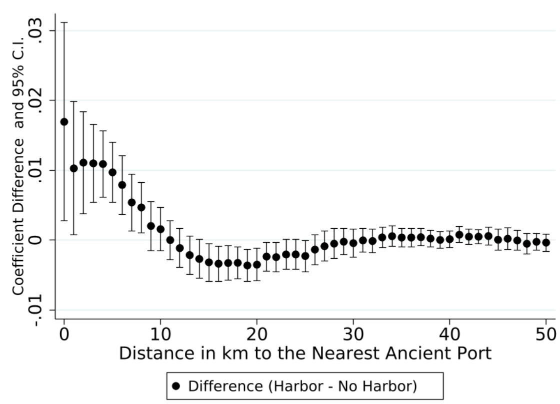
Panel A. Probability of City Density



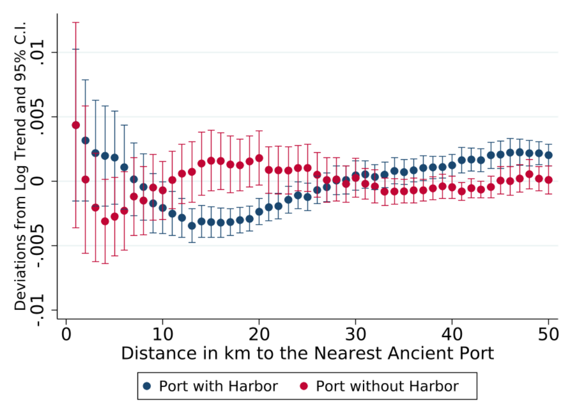
Panel B. Probability of City Density



Panel C. Differences for City Density



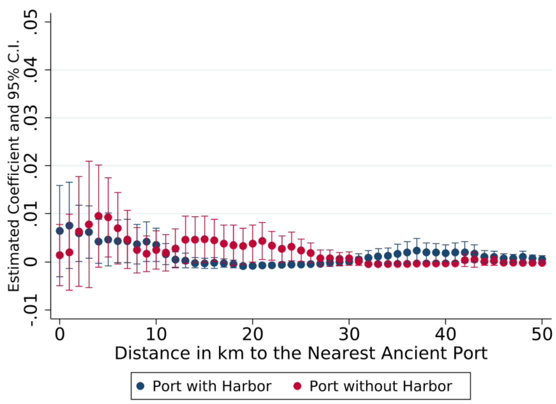
Panel D. City, Difference from Log Fit



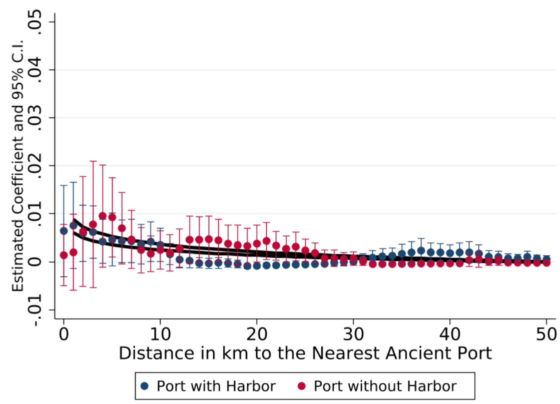
Notes: This figure reproduces estimates from Figures 6 and 7, restricting the sample to only grid cells that fall within the Roman Empire. We use the maps of the Roman Empire for 117 CE produced by the Digital Atlas of Roman and Medieval Cultures (DARMC), by McCormick and Polk (2017), which reflect the Roman Empire's maximal geographic extent excluding brief territorial holdings at the time in Armenia, Assyria, and Mesopotamia, and later in Caledonia. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.6. Heterogeneous Impacts on Probability of City Density: Only Outside the Roman Empire

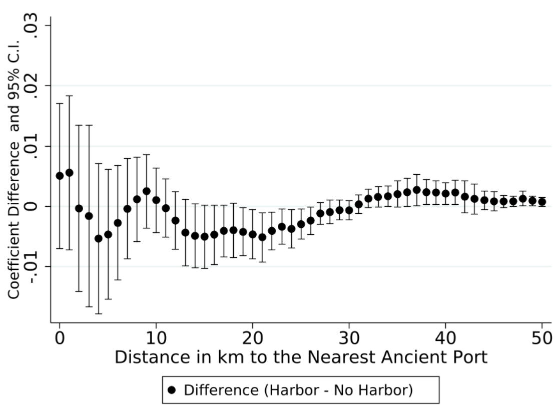
Panel A. Probability of City Density



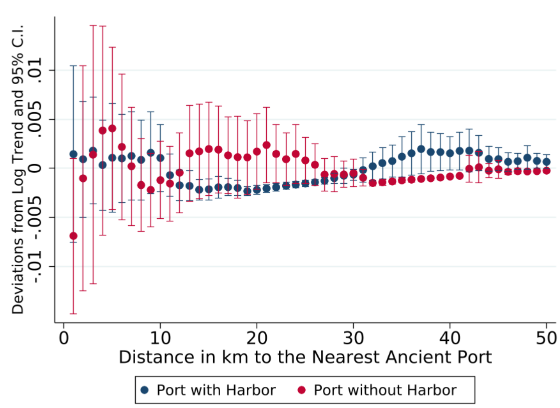
Panel B. Probability of City Density



Panel C. Differences for City Density



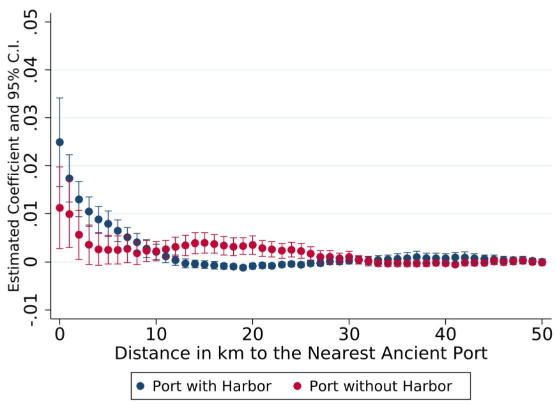
Panel D. City, Difference from Log Fit



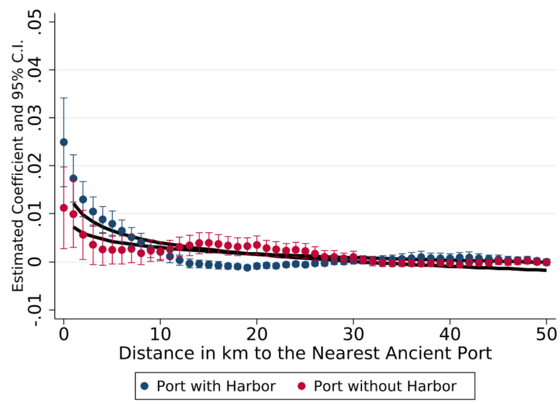
Notes: This figure reproduces estimates from Figures 6 and 7, restricting the sample to only grid cells that fall outside the Roman Empire (as defined in Appendix Figure A.5. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.7. Impacts on Probability of City Density, Robustness to Additional Geographic Characteristics

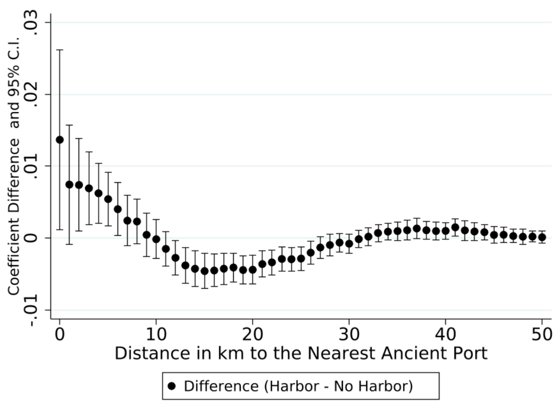
Panel A. Probability of City Density



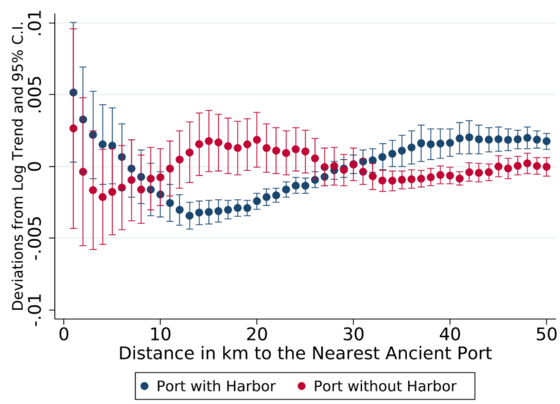
Panel B. Probability of City Density



Panel C. Differences for City Density



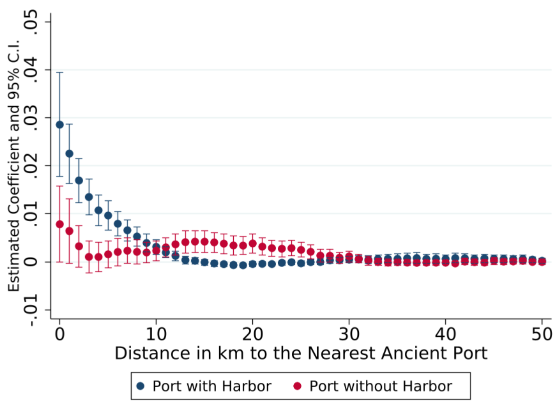
Panel D. City, Difference from Log Fit



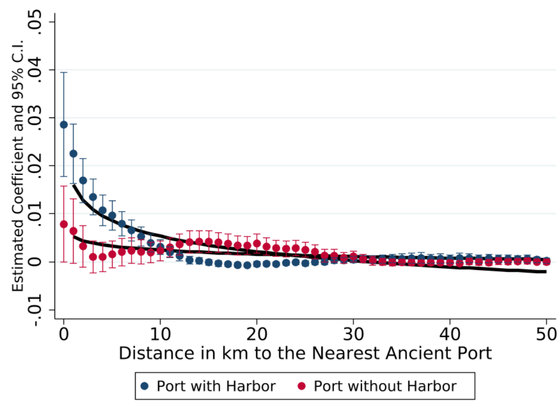
Notes: This figure reproduces estimates from Figures 6 and 7 with additional geographic controls for grid cell characteristics: suitability for potatoes, wheat, barley, cowpeas, olives, and chickpeas; sea bounty; an indicator for desert cell; an indicator for being on an island; elevation; log distance to river mouth; sea bounty interacted within being 2km, 5km, and 10km of the coast; indicators for being within 2km, 5km, and 10km of a coast, a major river, and a river mouth. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.8. Impacts on Probability of City Density, Robustness to Excluding Islands

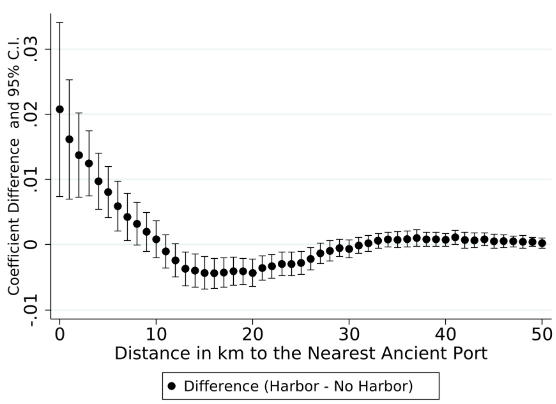
Panel A. Probability of City Density



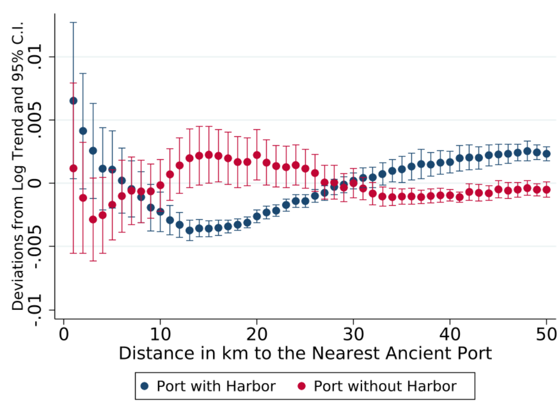
Panel B. Probability of City Density



Panel C. Differences for City Density



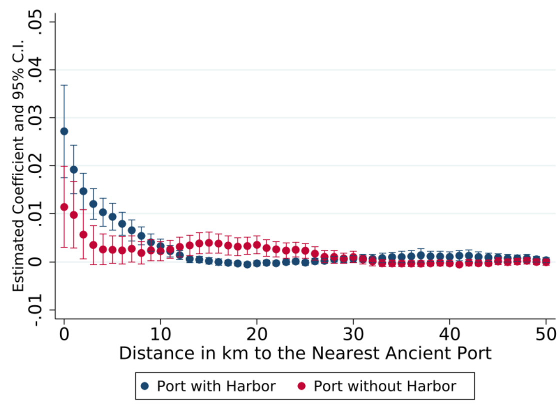
Panel D. City, Difference from Log Fit



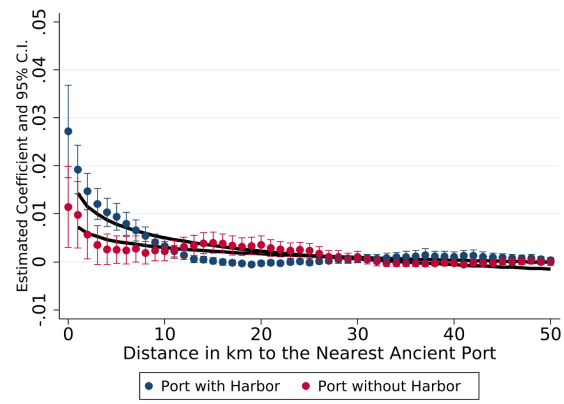
Notes: This figure reproduces estimates from Figures 6 and 7, removing all grid cells on an island smaller than 2,500 square kilometers. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.9. Impacts on Probability of City Density, Robustness to Excluding "Potential Ancient Harbors"

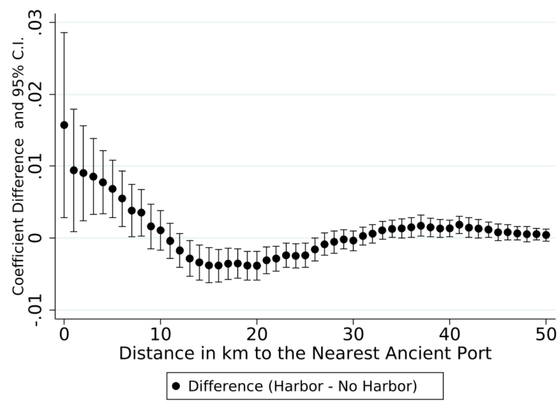
Panel A. Probability of City Density



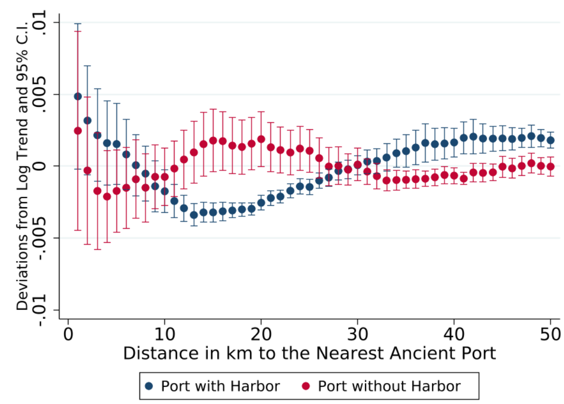
Panel B. Probability of City Density



Panel C. Differences for City Density



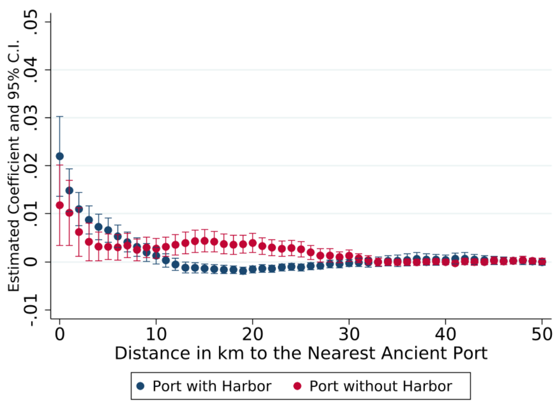
Panel D. City, Difference from Log Fit



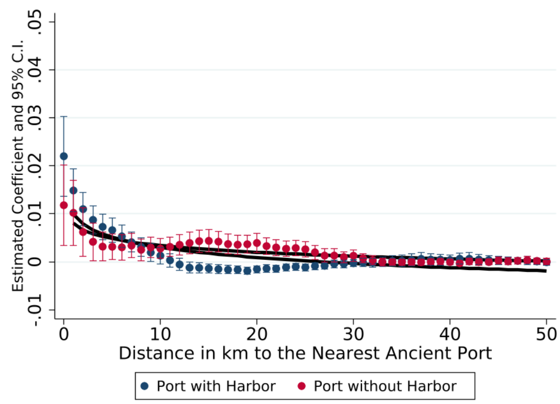
Notes: This figure reproduces estimates from Figures 6 and 7, excluding 216 locations where de Graauw is less certain of their use as ancient ports. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.10. Impacts on Probability of City Density, Robustness to Ancient Economic Characteristics

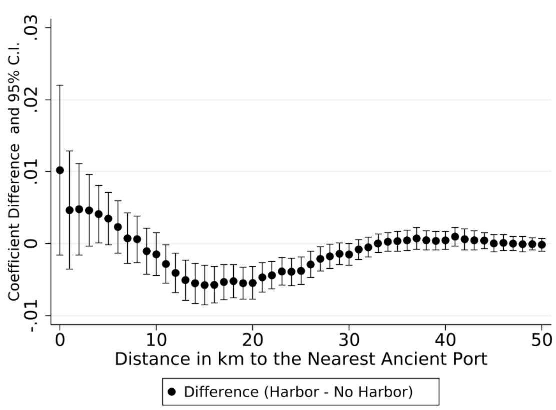
Panel A. Probability of City Density



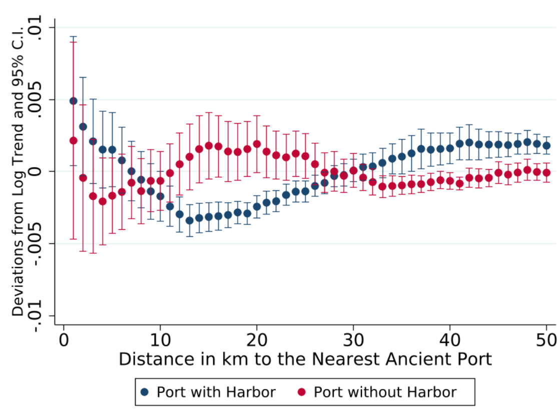
Panel B. Probability of City Density



Panel C. Differences for City Density



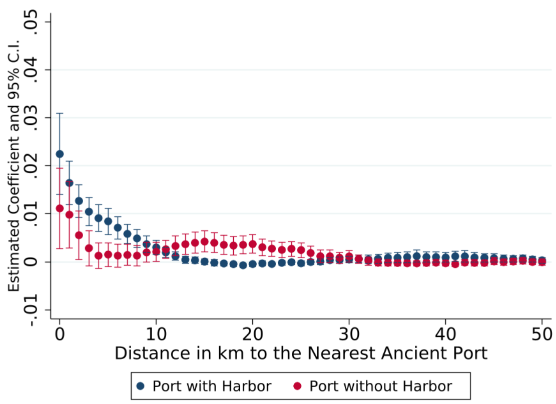
Panel D. City, Difference from Log Fit



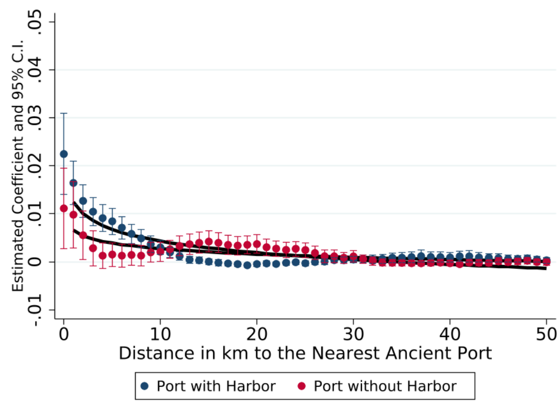
Notes: This figure reproduces estimates from Figures 6 and 7, with additional controls for grid cell characteristics in the ancient era: log distance to nearest Roman road and log distance to each Barrington 1 city (Talbert, 2000; Hanson, 2016): Aelia Capitolina (Jerusalem); Alexandria; Antioch (Antakya); Athens; Byzantium (Istanbul); Carthage; Córdoba; Corinth; Leptis Magna; Lugdunum (Lyon); Mediolanum (Milan); Rome; Tarraco (Tarragona); and Thessalonica. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.11. Impacts on Probability of City Density, Robustness to Excluding Ancient Open-Water Ports

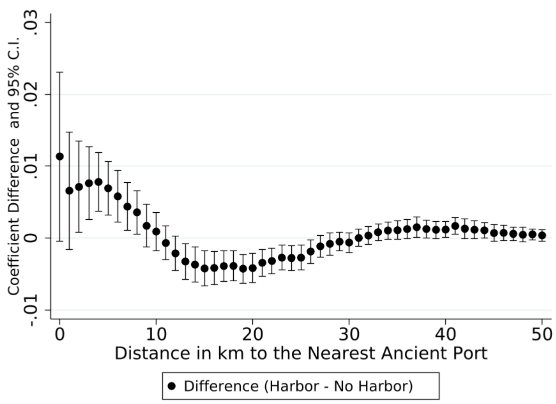
Panel A. Probability of City Density



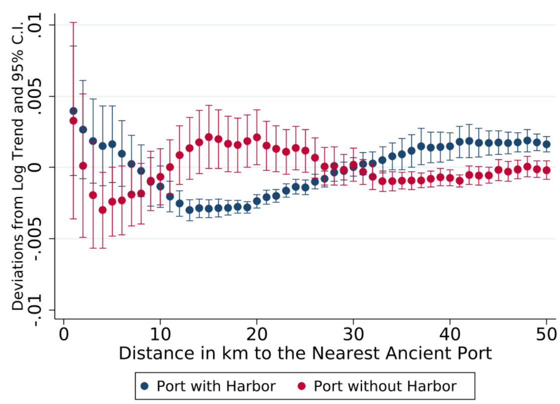
Panel B. Probability of City Density



Panel C. Differences for City Density



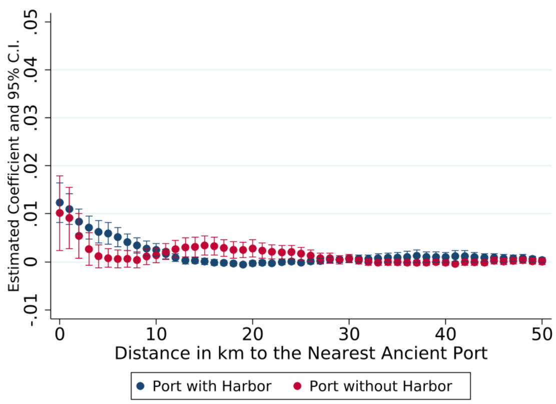
Panel D. City, Difference from Log Fit



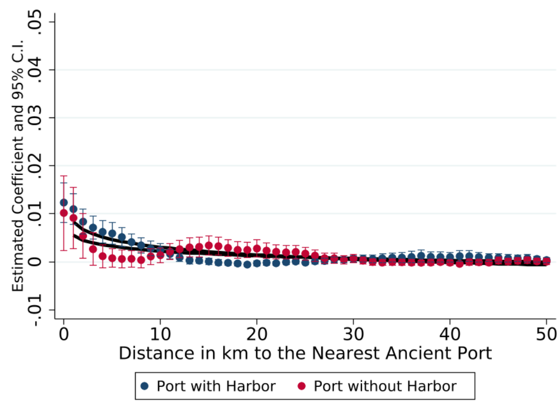
Notes: This figure reproduces estimates from Figures 6 and 7, removing all grid cells within 20km of 22 ancient open-water ports (those that are known to have relied on human-made protections from the sea). Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.12. Impacts on Probability of City Density, Robustness to Excluding Barrington 1 Cities

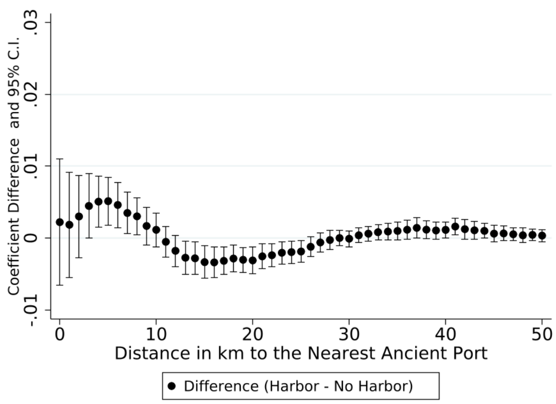
Panel A. Probability of City Density



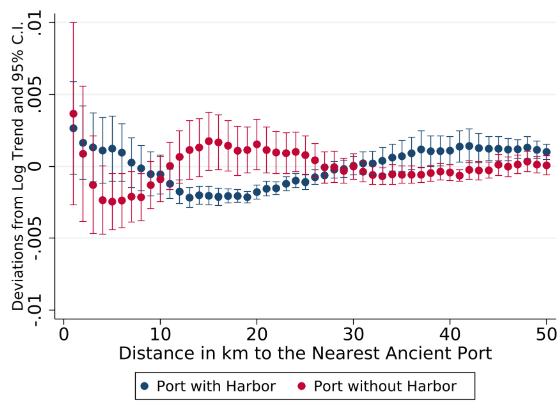
Panel B. Probability of City Density



Panel C. Differences for City Density



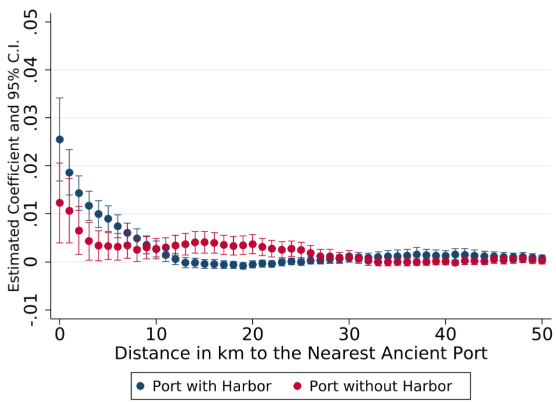
Panel D. City, Difference from Log Fit



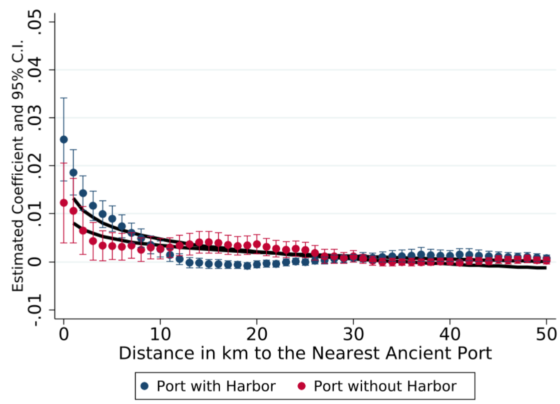
Notes: This figure reproduces estimates from Figures 6 and 7, removing all grid cells within 20km of 14 Barrington 1 Cities (Talbert, 2000; Hanson, 2016): Aelia Capitolina (Jerusalem); Alexandria; Antioch (Antakya); Athens; Byzantium (Istanbul); Carthage; Córdoba; Corinth; Leptis Magna; Lugdunum (Lyon); Mediolanum (Milan); Rome; Tarraco (Tarragona); and Thessalonica. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.13. Impacts on Probability of City Density, Robustness to 2-Degree by 2-Degree Fixed Effects

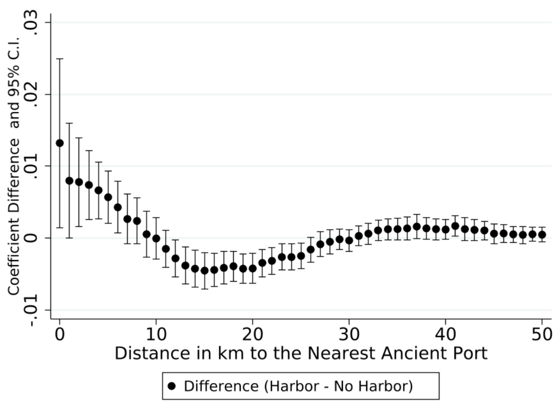
Panel A. Probability of City Density



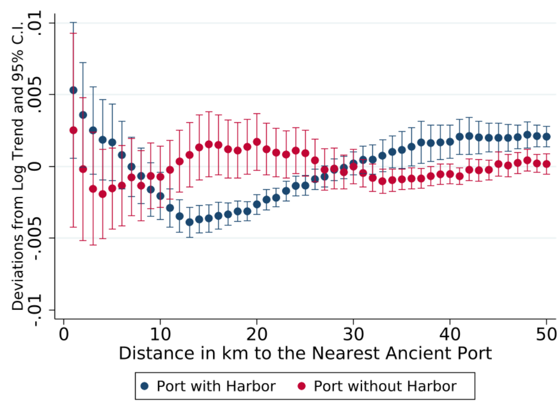
Panel B. Probability of City Density



Panel C. Differences for City Density



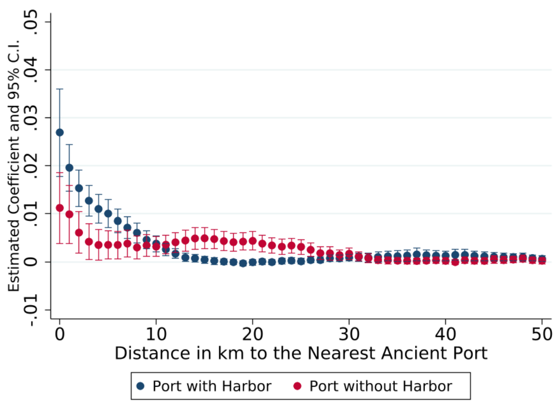
Panel D. City, Difference from Log Fit



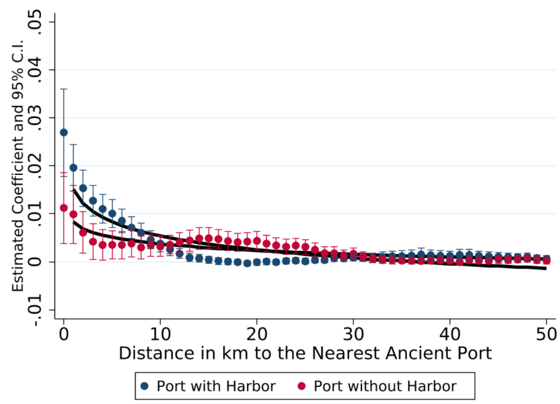
Notes: This figure reproduces estimates from Figures 6 and 7, adding fixed effects for 2-degree by 2-degree grid cell groupings. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.14. Impacts on Probability of City Density, Robustness to Modern Country Fixed Effects

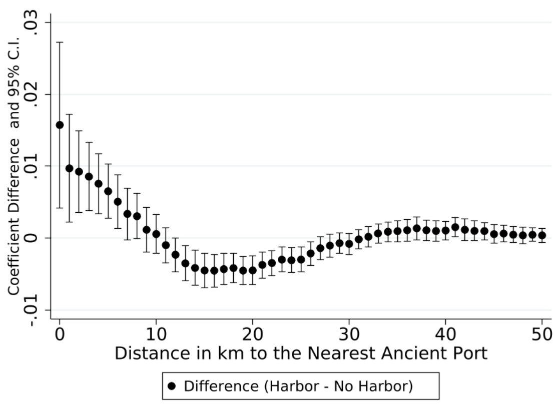
Panel A. Probability of City Density



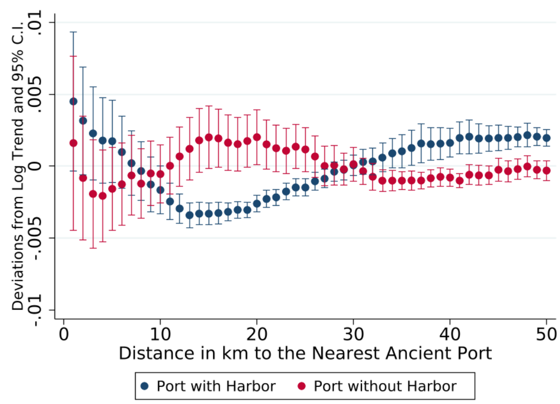
Panel B. Probability of City Density



Panel C. Differences for City Density



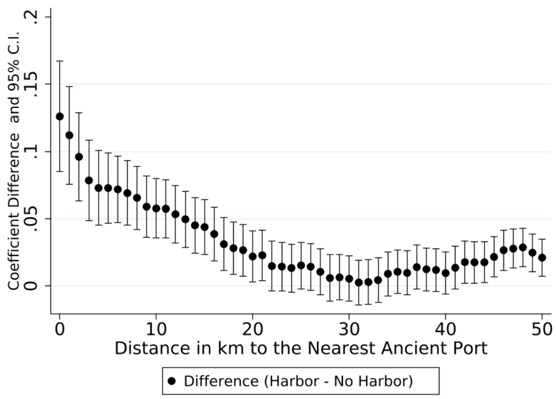
Panel D. City, Difference from Log Fit



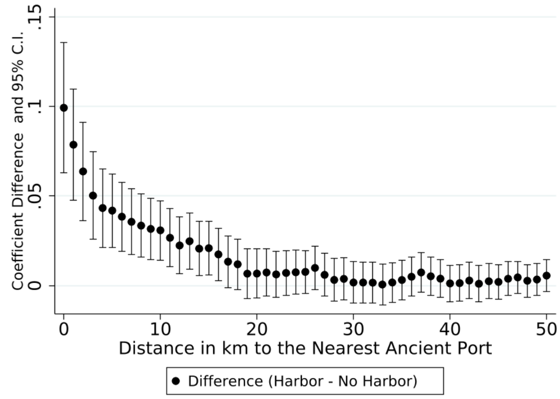
Notes: This figure reproduces estimates from Figures 6 and 7, adding fixed effects for present-day country. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.15. Coefficient Differences, Across Population Density Thresholds

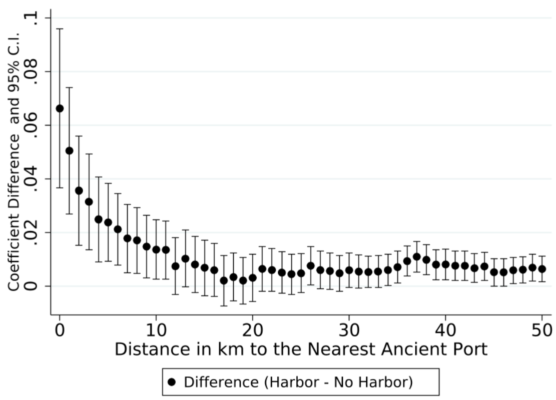
Panel A. Ln(Density) > 5



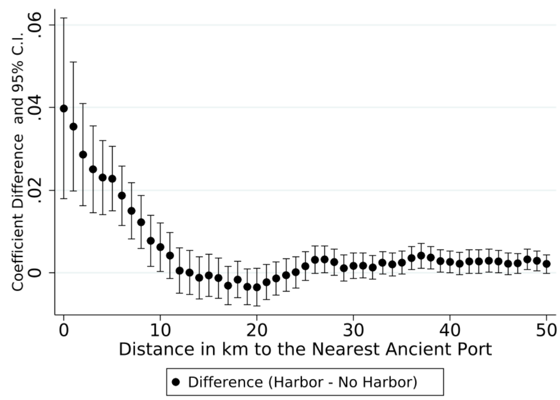
Panel B. Ln(Density) > 6



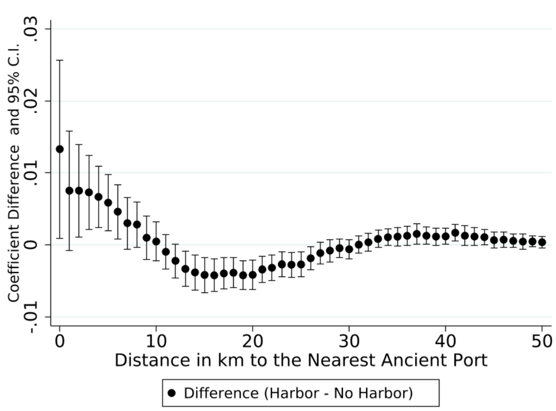
Panel C. Ln(Density) > 7



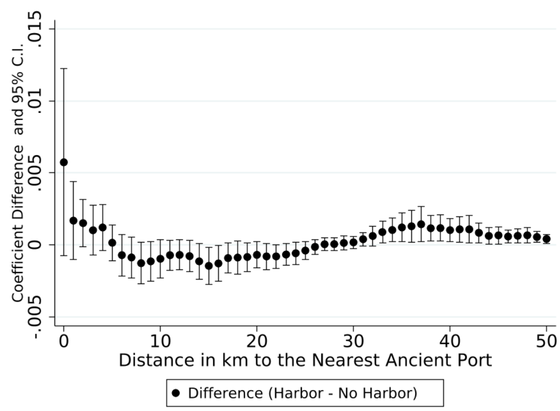
Panel D. Ln(Density) > 8



Panel E. Ln(Density) > 9



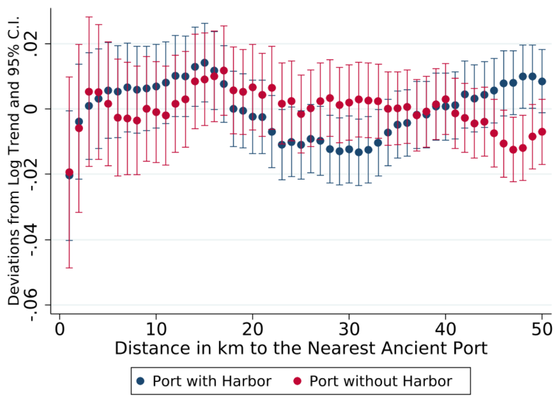
Panel F. Ln(Density) > 10



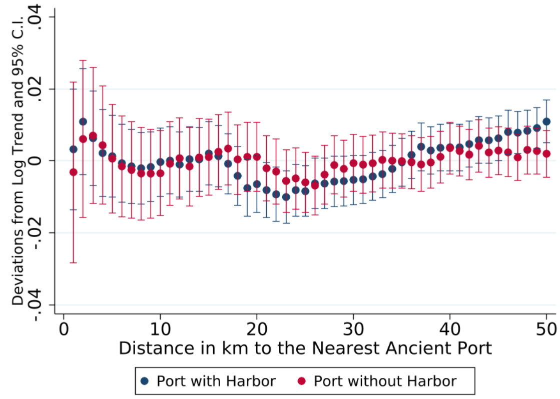
Notes: This figure shows the differences between the harbor and no harbor estimates of the effect of distance on different log population density thresholds. Panels B and E match Panels C and D of Figure 6, and the other panels report results for alternate log population density cutoff values from 5 to 10.

Figure A.16. Differences from Log Fit, Across Population Density Thresholds

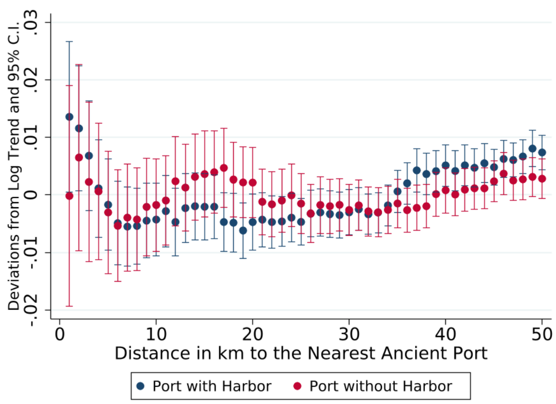
Panel A. Ln(Density) > 5



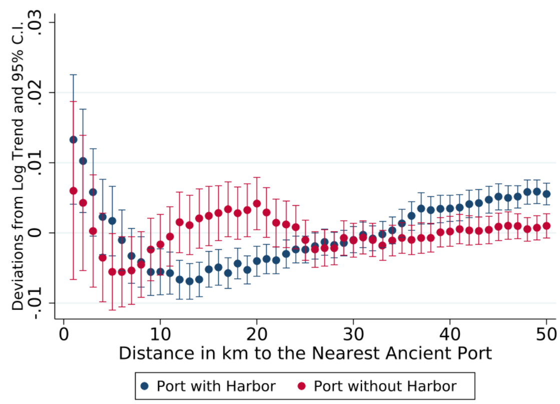
Panel B. Ln(Density) > 6



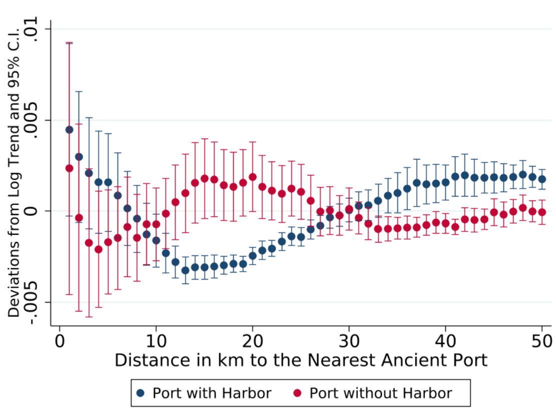
Panel C. Ln(Density) > 7



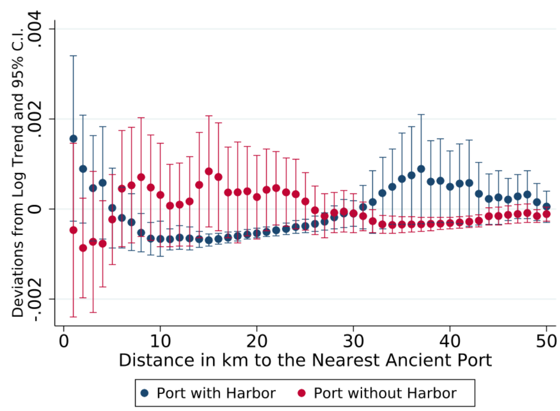
Panel D. Ln(Density) > 8



Panel E. Ln(Density) > 9



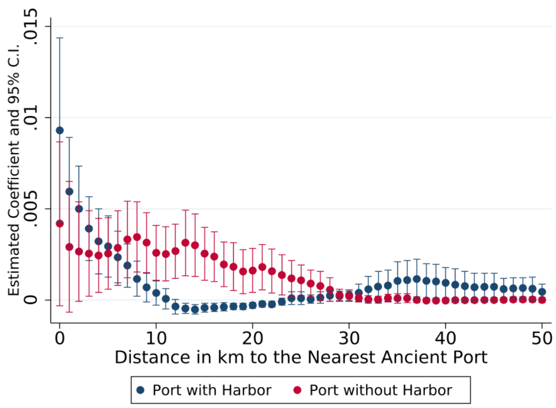
Panel F. Ln(Density) > 10



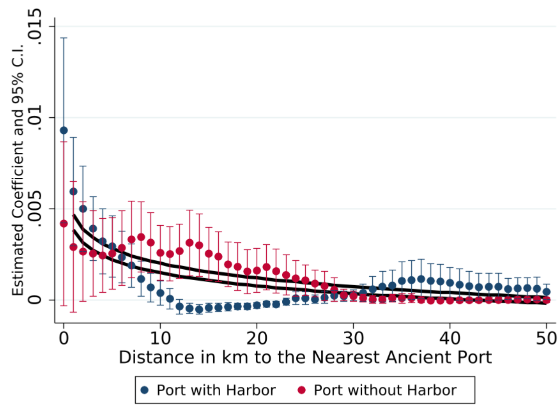
Notes: This figure shows differences between the log fit and the estimated effects of distance on different log population density levels, for distances to ports both with and without modern harbors. Panels B and E match Panels C and D of Figure 7, and the other panels report results for alternate log population density cutoff values from 5 to 10.

Figure A.17. Impacts on Probability of City Density, Robustness to 2000 GRUMP Data

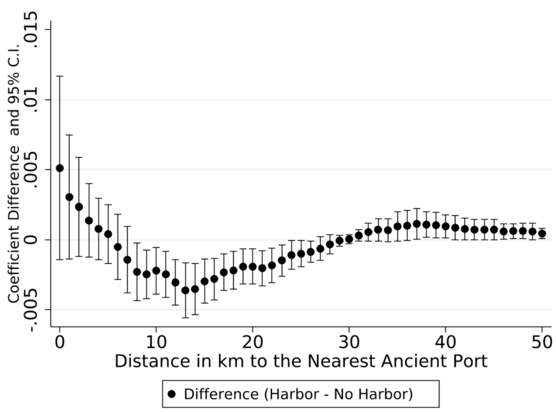
Panel A. Probability of City Density



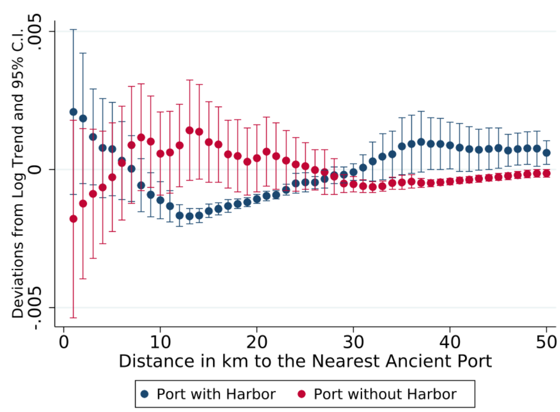
Panel B. Probability of City Density



Panel C. Differences for City Density



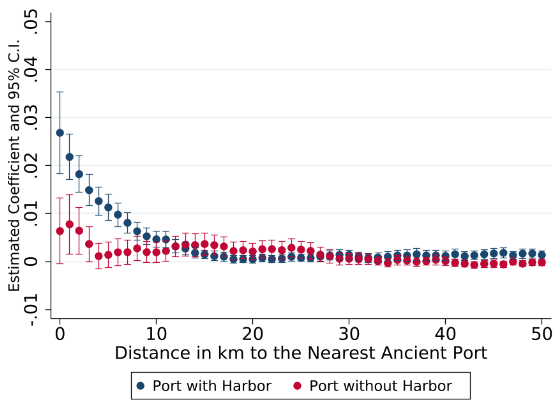
Panel D. City, Difference from Log Fit



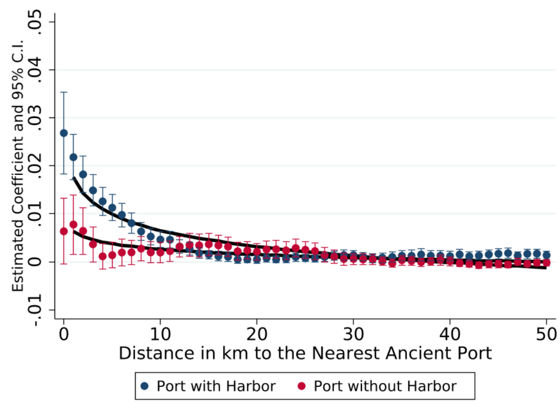
Notes: This figure reproduces estimates from Figures 6 and 7, using population per square kilometer estimates for 2000 from the earlier GRUMP model (CIESIN, 2011). Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.18. Impacts on Probability of City Density, Robustness to GHSL Density Data

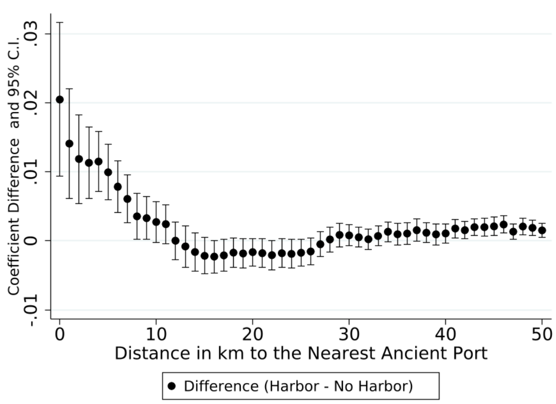
Panel A. Probability of City Density



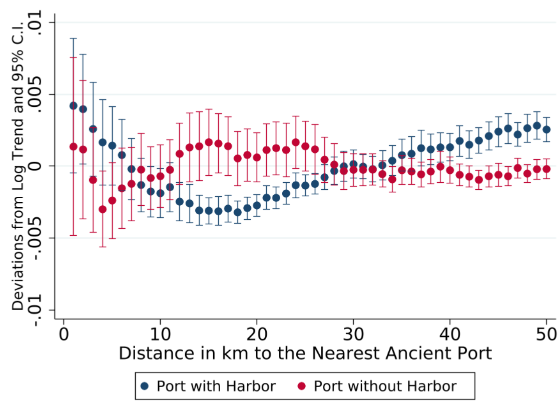
Panel B. Probability of City Density



Panel C. Differences for City Density



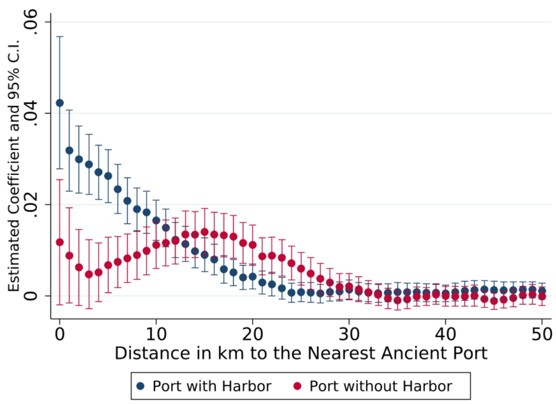
Panel D. City, Difference from Log Fit



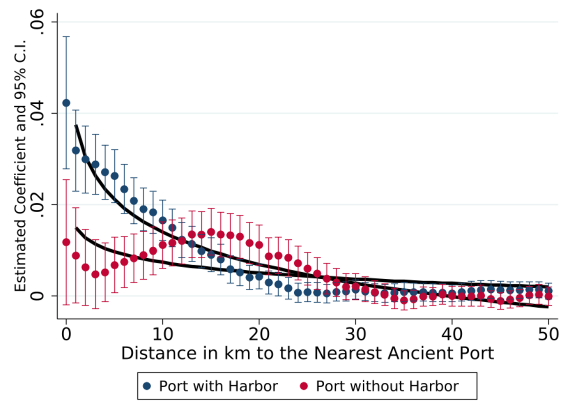
Notes: This figure reproduces estimates from Figures 6 and 7, using population per square kilometer estimates for 2015 from the GHSL-POP model (Schiavina, Freire, and MacManus, 2019). Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.19. Impacts on Probability of City Density, Robustness to GHSL City Location Data

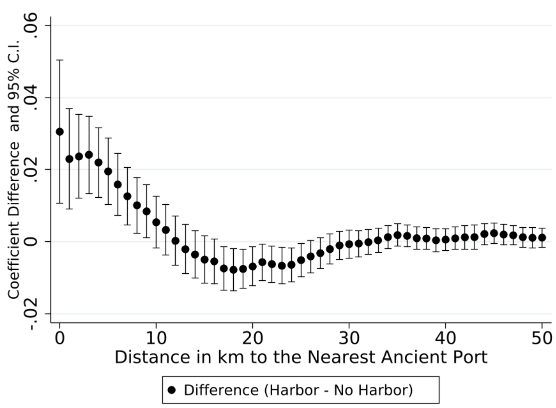
Panel A. Probability of City Density



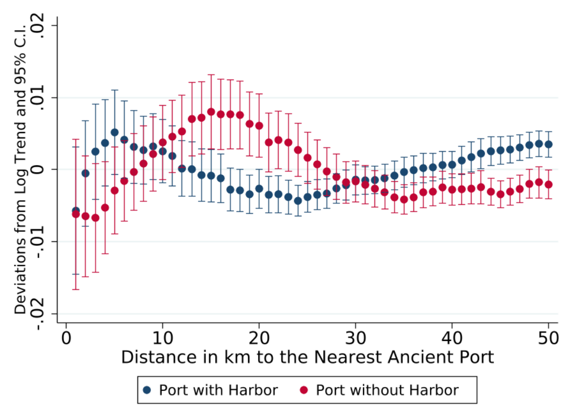
Panel B. Probability of City Density



Panel C. Differences for City Density



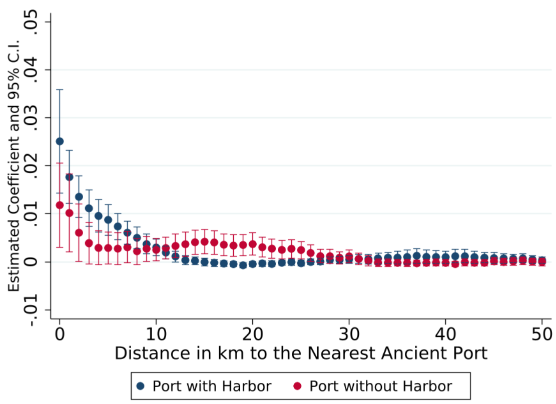
Panel D. City, Difference from Log Fit



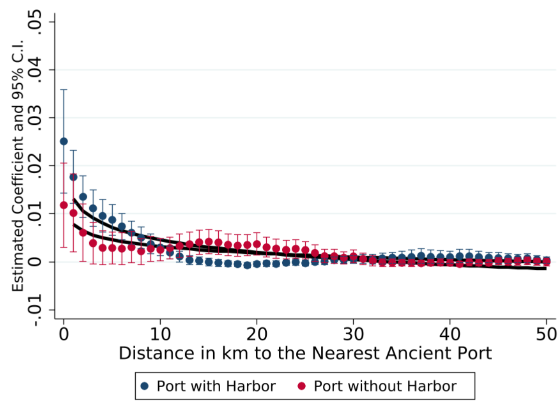
Notes: This figure reproduces estimates from Figures 6 and 7, using an alternative outcome of being within a city. We define a grid cell as within a city if it is within the "radius" ($\sqrt{\text{Area}/\pi}$) of a city center location, for cities over 500,000 people, using 2015 area and population estimates from the GHS Urban Centre Database (Florczyk et al., 2019). Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.20. Impacts on Probability of City Density, Robustness to Two-Way Clustered Standard Errors with Shifted Clusters

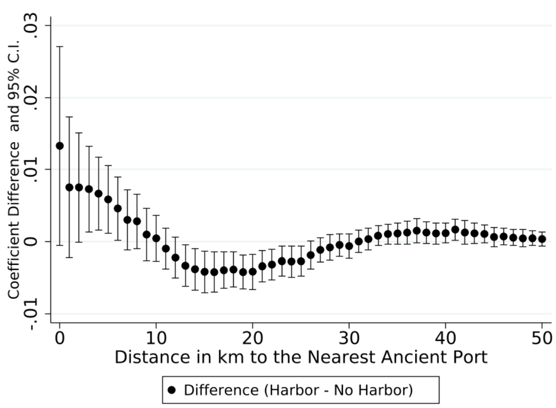
Panel A. Probability of City Density



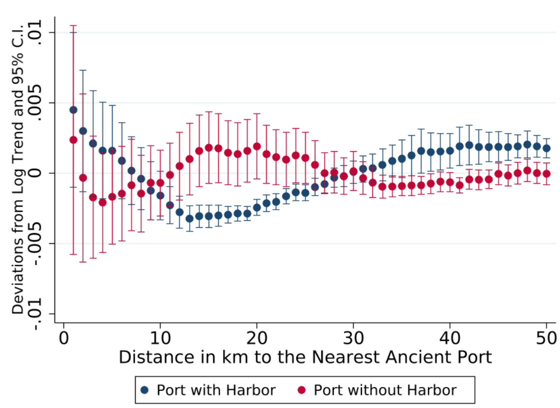
Panel B. Probability of City Density



Panel C. Differences for City Density



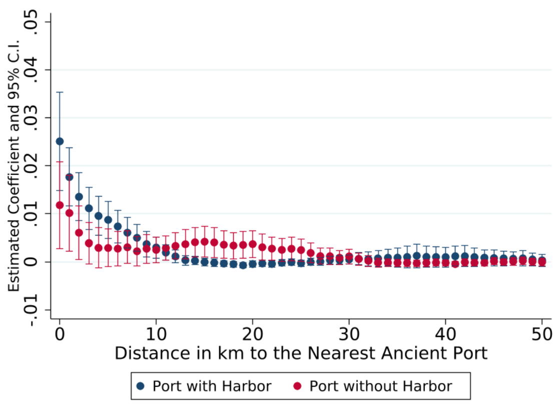
Panel D. City, Difference from Log Fit



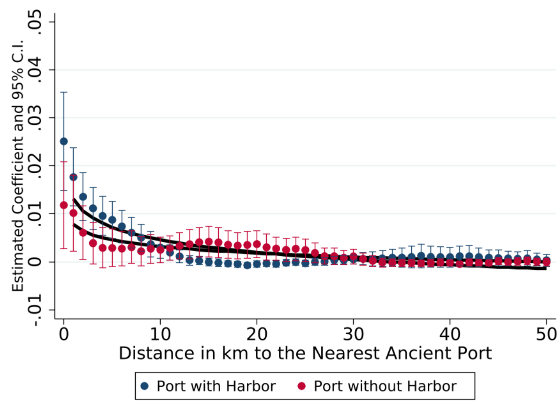
Notes: This figure reproduces estimates from Figures 6 and 7, with the estimation of standard errors using two-way clusters with additional clusters shifted by 1/24 degree latitude and 1/24 degree longitude from baseline. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.21. Impacts on Probability of City Density, Robustness to Clustering by 1/4-Degree by 1/4-Degree Squares

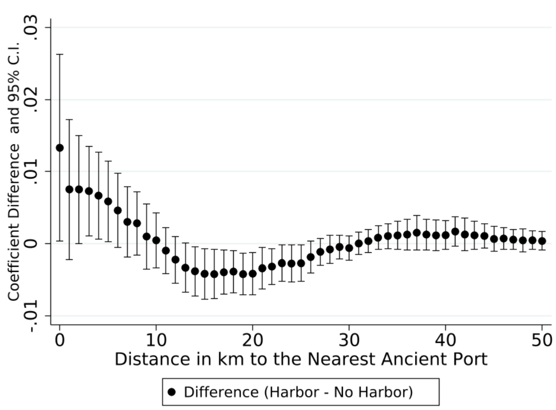
Panel A. Probability of City Density



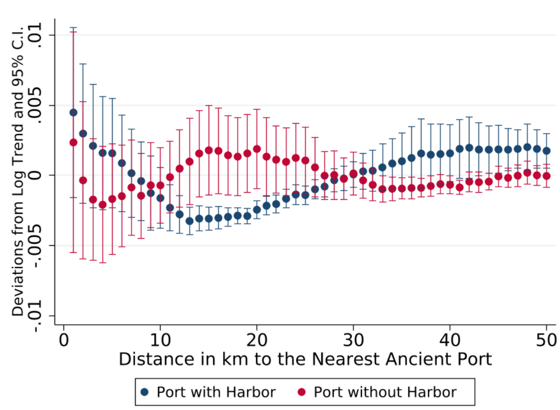
Panel B. Probability of City Density



Panel C. Differences for City Density



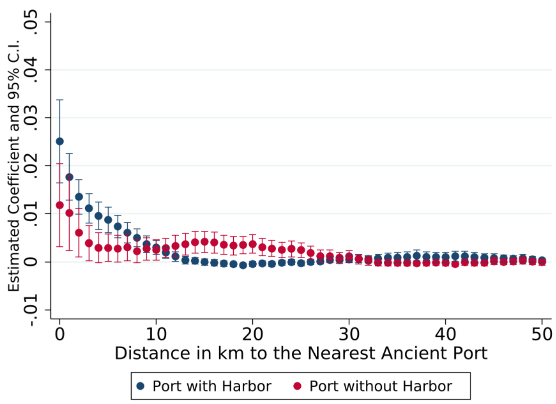
Panel D. City, Difference from Log Fit



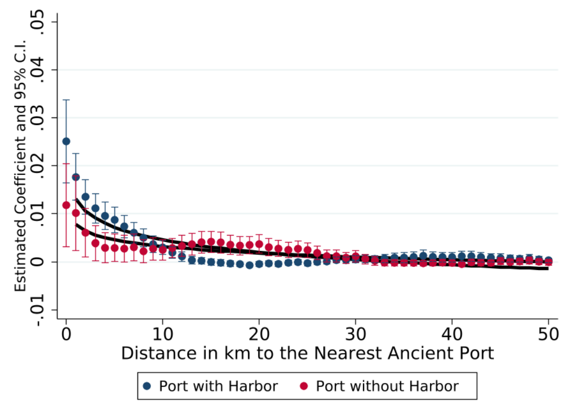
Notes: This figure reproduces estimates from Figures 6 and 7, with standard errors clustered instead at the level of 1/4-degree-by-1/4-degree groups. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.22. Impacts on Probability of City Density, Robustness to Using Conley Standard Errors with 4km Cutoff

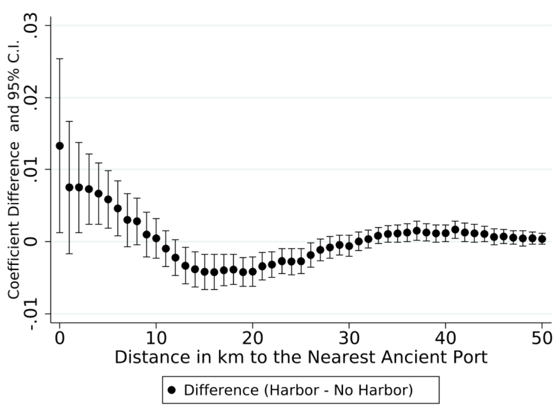
Panel A. Probability of City Density



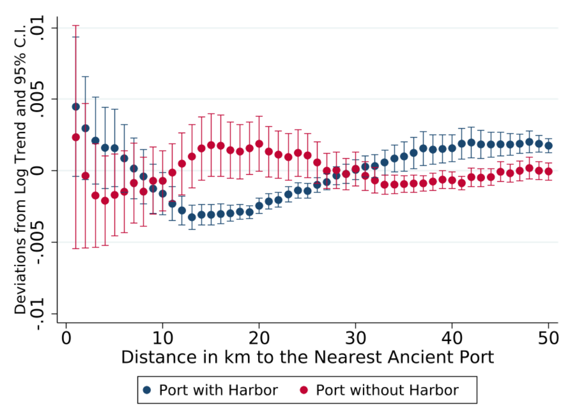
Panel B. Probability of City Density



Panel C. Differences for City Density



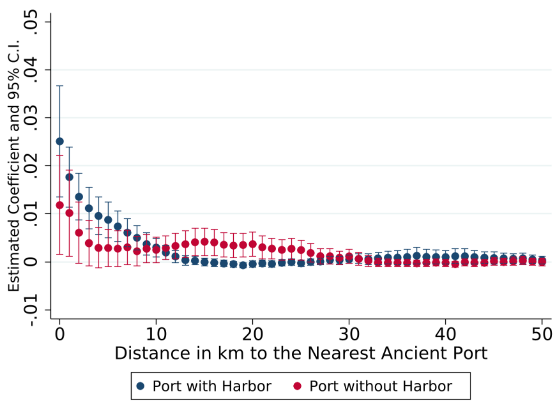
Panel D. City, Difference from Log Fit



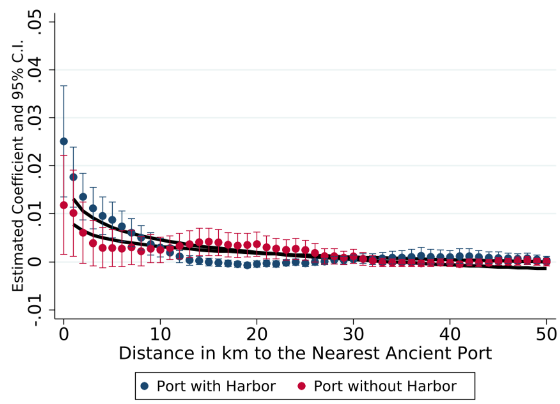
Notes: This figure reproduces estimates from Figures 6 and 7, instead using Conley standard errors (Conley, 1999; Colella et al., 2019) that allow for arbitrary spatial correlation within a 4km radius. By comparison, the median size city in our sample has a radius of 4km. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.

Figure A.23. Impacts on Probability of City Density, Robustness to Using Conley Standard Errors with 8km Cutoff

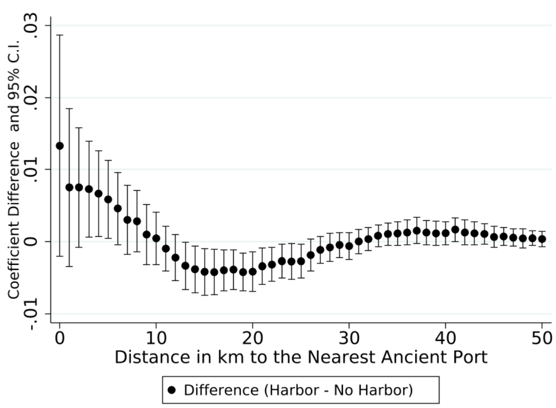
Panel A. Probability of City Density



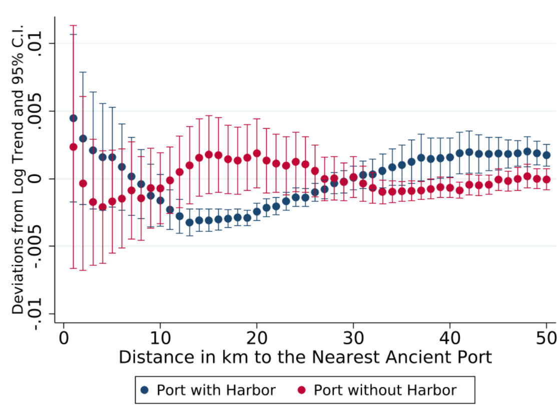
Panel B. Probability of City Density



Panel C. Differences for City Density



Panel D. City, Difference from Log Fit



Notes: This figure reproduces estimates from Figures 6 and 7, instead using Conley standard errors (Conley, 1999; Colella et al., 2019) that allow for arbitrary spatial correlation within an 8km radius. By comparison, the median size city in our sample has a radius of 4km. Panels A and C correspond to Figure 6, Panels B and D. Panels B and D correspond to Figure 7, Panels B and D.