ASYMMETRIC TRADING COSTS AND ANCIENT GREEK CITIES[1]

Yuxian Chen* Yannis M. Ioannides Ferdinand Rauch*****

September 14, 2026

Disclosures: The authors have nothing to disclose. Only resources of our universities, Tufts, Brown, Oxford, Heidelberg and St. Gallen were used in this research.

JEL Codes: F10, N70, N90, R12.

Keywords: ancient urbanization, intercity trade, asymmetric trading costs, intercity travel times, economic geography, spatial models, transportation networks, network centrality, systems of cities, market access, Hellenic poleis, ancient Greek cities, ancient Hellenic world.

*Department of Economics, Brown University, USA. yuxian_chen@brown.edu.

**Professor of Economics, Tufts University. Yannis.Ioannides@tufts.edu. Corresponding author. Address: Department of Economics, 177 College Avenue, Tufts University, Medford, MA 02155, USA.

***Professor of Economics, University of St Gallen, Switzerland. ferdinand.rauch@unisg.ch.

[1]We acknowledge very helpful comments by the editor, Theodore Papageorgiou, and two referees that allowed us to improve the paper. We also benefited from comments by a number of scholars but do not implicate them. They are: Jim Anderson, Costas Arkolakis, Yann Bramoullé, Alain Bresson, Kerem Coşar, Aureo De Paula, Yannis Evrigenis, Bruce Hitchner, Roland Hodler, Misha Kilmer, John Kroll, Hon Ho Kwok, Jacques Melitz, James Murphy, David Nagy, Josiah Ober, Theodore Papageorgiou, Herakles Polemarchakis, Harilaos Psaraftis, Diego Puga, Edouard Schaal, Peter Van Alfen, Conor Walsh and seminar participants at various institutions. We also acknowledge help from Stelios Michalopoulos, who contributed to the original construction of the data on sailing times among poleis, and from Nick Reynolds, who actually constructed the original sailing costs dataset. We are grateful to Dr. Takvor Sukissian of the Hellenic Center for Marine Research, who helped with updating the data that Christopher Barnett, Carolyn Talmadge, Uku-Kaspar Uustalu and the other members of the Tufts Data Lab team on sailing times across the coastal segments of the Mediterranean and the Black Seas had helped us with. We benefited from excellent research assistance by Joseph Chance, Xu Dong, Gengbin Piao and James Plosser.

ASYMMETRIC TRADING COSTS AND ANCIENT GREEK CITIES

Yuxian Chen* **Yannis M. Ioannides**** **Ferdinand Rauch*****

Abstract

Asymmetric trading costs arise when shipping times from point i to point j differ from shipping from point j to i. We show that such asymmetric trading costs predict distinct patterns of location in a class of models using Anderson-Dixit-Stiglitz preferences, where advantage in import market access (or potential) is more beneficial than that in export market access. We bring this prediction to the data, as well as additional ones along the lines of a number of propositions, while also controlling for local micro-geographic characteristics and using data on the location and sizes of cities in ancient Greece. Prevailing winds create an environment of asymmetric and seasonally varying trading costs in ancient Greece.

We find that trade potential is strongly associated with the location and size of poleis. The model's directional prediction is supported when trade potential is measured using the left and right eigenvector centralities implied by the model: import-oriented centrality has a larger coefficient than export-oriented centrality. The result is less robust for alternative reduced-form measures. Using 2SLS methods for population-weighted market access, we find significant differences between export and import market access, with the former often more important than the latter. These findings show that directional asymmetries in trading costs matter for urbanization, while also suggesting that their empirical implications depend on how directional access is measured.

*Department of Economics, Brown University, USA. yuxian_chen@brown.edu.

**Professor of Economics, Tufts University. Yannis.Ioannides@tufts.edu. Corresponding author. Address: Department of Economics, 177 College Avenue, Tufts University, Medford, MA 02155, USA.

***Professor of Economics, University of St Gallen, Switzerland. ferdinand.rauch@unisg.ch.

1 Introduction

Most of the trade literature assumes that the cost of trade from location i to location j is equal to the cost of trade from location j to i. Yet there are examples where this assumption does not hold. Along a river, it is usually faster to ship downstream than upstream. Transport costs of a container often differ widely by direction of trade. Congestion creates asymmetries by travel direction and time. Prior to steamship, strong wind patterns allowed ships to sail cheaply in one direction, yet return in a circular fashion. In this paper, we study this phenomenon of asymmetric trade costs.

We make two main contributions. First, we show theoretically that asymmetric trade costs affect the attractiveness of locations in a standard Anderson-Dixit-Stiglitz framework once the conventional symmetry assumption is relaxed.[2] Second, we test the model's predictions using data on the locations of cities in Hellas, which encompasses not only ancient Greece but also what the classics literature refers to as the ancient Hellenic world. This setting is particularly well suited to studying asymmetric trade costs because trade relied heavily on sailing technology and was therefore shaped by strong wind patterns. We present evidence that these asymmetries had non-trivial effects on the locations of economic activity. As part of this analysis, we also document new facts about the locational fundamentals of ancient Greek cities, including the role of seasonal variation in trading costs, proxied by travel times.

Economic, political, and institutional activity in ancient Greece was centered in its cities—more precisely, its city-states, or *poleis* [Hansen (2006)].[3] Classicists and historians have documented extraordinary patterns of trade, growth, and institutional and political innovations. Influential works by Ober (2015) and Bresson (2016) have emphasized the extraordinary growth that was followed by a secular decline in the classical Greek era in the context of primitive but nonetheless vibrant market economies. Much of the present research is based on data from the record of archaeological finds and text references, compiled by the Copenhagen Polis Centre and then reported and digitized by Hansen and Nielsen (2004).[4]

[2] We use Anderson-Dixit-Stiglitz to label the large literature that has used CES preferences in trade models.

[3] The paper follows the convention in the classics literature and adopts Greek spelling, so that poleis is the plural for *polis*. Relatedly, the main body of the paper should refer to poleis, but in deference to the modern reader, we refer to cities. In the ancient Greek context, polis also refers to a territory, which includes concentrated population centers, settlements, that may be referred to as cities. Still, polis is a broader concept that refers also to their respective population and its political and social organization. The meaning of polis did not change significantly over the archaic and classical period, which the latter by convention ends at 423 BCE; Hansen and Nielsen (2004), Ma (2024).

[4] These data were digitized originally by Josiah Ober and his team of Stanford scholars and were augmented

Those together with additional data from other sources are amenable, in principle, to quantitative analysis. The literature has emphasized that the broader Greek world, and especially the larger poleis, the experienced growth of population and improvement in living standards after the Hellenic Dark Age (Scheidel 2003).

Our analysis focuses on the geographic and network determinants of city location in the ancient Hellenic world. We ask which geographic advantages help explain the locations of ancient settlements, including both local natural fundamentals and trade potential, by which we mean a site's access to other locations through the trading network, and how seasonal variation in travel times affects the attractiveness of potential locations.[5]

We address these questions through an exhaustive analysis of the poleis for which location and size information are available. We compare their observed locations with the universe of potential sites along the Mediterranean and Black Sea coastlines. Using our expanded dataset for all coastal segments, we examine whether the locations at which cities emerged can be explained by local geographic fundamentals, access to the broader trading network, and directional and seasonal variation in travel costs.

1.1 Summary of Findings

Our empirical results show that trade potential was an important determinant of the location and size of poleis in the ancient Hellenic world. We use two measures of trade access. The first is network centrality, derived from our theoretical model of interpolis trade, which builds on a standard model of international trade and emphasizes asymmetric trading costs.[6] The second is market-access measures, conceptualized as outward (or export), and inward (or import) access (or potential), commonly used in empirical work. Both network centrality and market access predict settlement outcomes, and the directional asymmetries in shipping costs play an important role in these relationships.

We first examine trade potential relative to the universe of possible settlement sites. We divide the Mediterranean and Black Sea coastlines into evenly spaced segments and use the

by ourselves working from Hansen and Nielsen (2004).

[5]Hicks (1969) argued that Mediterranean geography was conducive to urban development in the broadest economic, intellectual, and political senses, long before such issues were addressed quantitatively. More recently, Abulafia (2011) emphasized the significance of the Mediterranean in facilitating human interaction, commerce, and cultural exchange over and above urban development. Matsuyama (2017) similarly highlights the importance of the proximity, or adjacency, matrix in amplifying the geographical advantages and disadvantages of regions and thereby facilitating inquiries in economic history *inter alia*.

[6]This framework is particularly appropriate in this setting because the ancient Greek city-states were effectively sovereign polities that shared many geographic and institutional characteristics.

midpoint of each segment as the unit of analysis. We then relate settlement outcomes to each segment's trade potential with respect to all other coastal segments. These regressions provide evidence that locations occupying more advantageous positions in the trading network were more likely to host poleis and larger settlements. They also show that accounting for the directional nature of trading costs materially affects the estimated importance of trade access.

The model predicts that import access (or potential) should matter more than export market access, a prediction supported in the specification closest to the theory. When trade potential is measured by the left and right eigenvectors implied by the model, the left eigenvector, which corresponds to import potential, has a larger coefficient than the right eigenvector, which corresponds to export potential.

This result is less robust across alternative measures of directional trade potential. In regressions using Kleinberg centralities or population-weighted market access to observed settlements, export access is generally more important than import access. This difference may reflect the fact that these measures are reduced-form proxies rather than the exact structural objects in the model. It may also reflect scale-dependent trade costs, under which export access can become relatively more important.

Because access to observed settlements is endogenous, we estimate these market-access specifications using two-stage least squares and several alternative instruments, which are described in detail in the empirical section. The estimates generally show statistically significant effects of access to poleis near other coastal segments, weighted by the relevant directional shipping costs.

Finally, we examine both seasonality and seasonal variation in shipping conditions. Using season-specific travel costs, we find that seasonal trade potential is systematically related to settlement size. In particular, access to valuable trading destinations during the summer appears to be more important for urban development than access during other seasons. At the same time, we find that greater seasonal variation in travel costs can be advantageous. One possible interpretation is that cities can time their voyages to take advantage of the season in which shipping costs are lowest.

1.2 Roadmap

The remainder of the paper proceeds as follows. Section 2 reviews the related literature, and Section 3 provides background on the economy of ancient Greece. Section 4 develops a model of trade among poleis under the Armington assumption, using the Anderson-Dixit-Stiglitz framework and the spatial-equilibrium approach of Allen and Arkolakis (2014, 2025) to develop a number of propositions and their empirical implications that emphasize asymmetric trading costs and pertain to the attractiveness and size of urban settlements. The model yields several propositions, of which two anchor the analysis. First, asymmetric trading costs generate distinct importer-side and exporter-side Perron–Frobenius eigenvectors, implying that inward and outward network positions need not have the same relationship with equilibrium populations, and that the magnitude of asymmetry matters. Also, the marginal value of reducing a directional trade friction is itself direction-specific, depending on the importance of the connected locations on the relevant importer and exporter sides of the network.

Section 5 describes the data. In particular, Sections 5.1–5.3 provide details on the urban settlements, or poleis, and the coastal segments, while Sections 5.4 and 5.5 develop the measures of trading costs and urbanization used in the empirical analysis. Section 6 presents the main empirical results. It first examines the importance of trade potential relative to the universe of possible settlement sites and then studies market access to observed settlements using instrumental-variables methods. Section 7 provides additional evidence on the effects of seasonality and seasonal variation in travel costs. Section 8 concludes.

The framework we propose below may also be extended to study how alliances, institutional diffusion, and political networks interacted with trade to shape the emergence and size of poleis. We plan to address these dimensions in future extensions of this work.[7]

2 Literature Review

The present paper restricts itself to a quantitative investigation of basic facts, which is rare in the humanities-oriented literature. Ober (2015) and Bresson (2016) do invoke broad aspects of economic theorizing in looking at the evidence, but do not adopt quantitative measures in any particular detail. Amemiya (2007) studies qualitatively the economy and economics of ancient Greece (with a special emphasis on the economy of Athens). A notable exception is Barjamovic *et al.* (2019), who use modern trade theory in conjunction with ancient commercial records data from Assyrian tablets to structurally estimate ancient city sizes and offer evidence in support of the hypothesis that large cities tend to emerge at the

[7] Issues of diffusion of technology in the form of alphabetic writing and coinage by poleis is also important but taken up in a separate investigation; see Chen and Ioannides (2026).

intersections of natural transport routes, as dictated by topography. They also seek to locate hitherto unidentified ancient Assyrian cities.

Next we review briefly a number of recent data-based studies that consider various aspects of Hellenic poleis in the light of modern tools. Chronopoulos *et al.* (2021) investigate the long-term effects of ancient colonialism on economic development. As it has been established by archaeologists, Phoenicians, Greeks and Etruscans spread around the Mediterranean, from the 11th to the 6th centuries BCE. By founding colonies, they transferred their superior technologies and institutions to local populations in new geographic areas with which they intermingled.[8] These authors find that geographic sites colonized by those civilizations are more relatively developed in the present day. Their results hold after controlling for contemporaneous country fixed effects and splitting the sample by continent. Moreover, their findings are robust to the use of alternative measures and different historical data sources on ancient colonies. Overall, those results suggest that ancient colonialism along the Mediterranean left a positive economic legacy which persists today despite two millennia of historical turbulence. They measure contemporary economic activity by means of the satellite-based lights data for 50 x 50 km grids [Henderson *et al.* (2018)]. They also employ the poleis data for robustness; see *ibid.* Table 6. Clearly, the results imply that ancient settlements are associated with superior locations, however in the setting those authors examine it is difficult to distinguish between persistent effects of first nature or path dependence in the first place.

Bonnier *et al.* (2020) report research where pollen data from a number of sites in southern Greece and Macedonia are used to study long-term vegetation change in these regions from 1000 BCE to 600 CE. They interpret, using insights from environmental history, observed trends in the regional presence of cereal, olive, and vine pollen as proxies for structural changes in agricultural production. They present evidence that there was a market economy in ancient Greece and a major trade expansion several centuries before the Roman conquest. They argue that their findings are consistent with auxiliary data on settlement dynamics, shipwrecks, and evidence of ancient oil and wine presses. The evidence reported by those authors on identification of cultivation types and dating via pollen as proxies for structural changes in agricultural production and trade is critical to their argument of the change in specialization; see *ibid.*, Figures 2 and 3.[9] Scholars have established that classical Athens

[8]For example, this explains the Phoenician origin of the Greek alphabet; relatedly, see Chen and Ioannides (2026). Similarly, colonists from Chalkis colonized Cumae in modern Italy via their own Aegean colony of Kymi in Euboia and passed on their own version of the Greek alphabet (script) to the Etruscans, through whom it found its way to the Romans and came to be known as the Latin alphabet.

[9]Specifically, Bonnier *et al. 2020*, 24-25, state that for southern Greece their estimated trends demonstrate

imported cereals from the Black Sea area [Bresson (2016)] in exchange for manufactures, while olive products were traded on a local or interregional scale.[10] In contrast, in Macedonia trends in the relative presence of cereals and olives go in the same direction and occurred in several phases before the Roman conquest in the 2nd century BCE. Such an empirical identification of market forces affecting trade across regions of Greece is supportive of reliance on economics tools in studying ancient urbanization.

Bakker *et al.* (2021), a study related to ours, report a causal effect of trade on economic development by appealing to the first systematic crossing of open seas in the Mediterranean during the time of the Phoenicians as the earliest massive trade expansions. They construct a measure of connectedness along the shores of the Mediterranean sea, which counts the number of other cells each coastal cell (with cells being defined as 10km $\times$ 10km spatial units) is connected with, given a maximum geographic distance. This connectivity varies with the shape of the coast, the location of islands, and the distance to the opposing shore. They relate connectedness to local growth, which they proxy via the presence of archaeological sites in an area using a number of alternative definitions. Better connected locations are associated with more archaeological sites during the Iron Age. The sensitivity of the coefficient of connectedness to varying its definition is surprisingly small. It is not clear whether this is a property of the geography. As the distance varies, the estimate peaks around 0.13 near 500km, but varies more dramatically when the time frame changes. In an alternative formulation, these authors use how many settlements a ship can reach within a given distance, arguably a more direct measure of market access. To account for the endogenous location of settlements they instrument this market access measure with the connectedness variable, both in logs. They corroborate their findings for the Mediterranean by expanding their study to the entire world while using a coarser definition of cells. Our study is closely related to Bakker *et al.* (2021) in terms of dataset structure, empirical approach and research questions. The main insights from our paper are not part of their analysis, in particular the importance of asymmetric trading costs as well as our findings regarding the importance of seasonal differences. We believe this is a crucial feature of the navigational environment in ancient

that despite the demographic growth visible in Figure 2, cereals played an increasingly smaller role in southern Greek agriculture and were becoming less and less visible in the landscape (Figure 3). At the same time, the importance of olive cultivation was steadily increasing. This prompts a natural question: why local producers chose to plant olives instead of sowing grains, when the demand for this basic foodstuff seemed to be increasing? Two possible explanations that are not mutually exclusive. First, the southern mainland of Greece was developing an export economy based on cash crops, such as olive cultivation, which would be consistent with the exploitation of the comparative advantage of this region.

[10] The increase in the relative importance of olive cultivation may have been due to increase in local demand by the elites of the Greek city-states as a consumption as well as luxury good associated with sports in the gymnasia.

times, and our paper offers a novel contribution.

The present paper emphasizes geography in the broad sense as a fundamental determinant of ancient Greek urbanization via trade among the poleis. It is thus abstracting from such political economy consideration as formation of alliances and their impact on trade. However, we do note that the classics literature emphasizes the significance of the institutional structure of the Delian League for trade, due to common coinage and the harmonization of standards for weights and measures via the imposition of the Athenian *Coinage, Weights and Measures Decree* [see Kallet and Kroll (2020)] on the group of allied poleis under the hegemony of Athens. [11] Our theoretical model readily predicts that increases in trading costs among some of the members of a trading system reduces welfare outcomes for all members, with piracy having similar effects.

Our paper also relates to an existing literature in trade that studies asymmetric trade costs. Anderson and van Wincoop (2003) do emphasize the importance of asymmetric trade costs, but in much of the trade literature the emphasis is on asymmetry generated by unequal tariffs at national borders. As Redding and Rossi-Hansberg (2017) conjecture, asymmetric shipping costs have, by breaking the proportionality that holds under symmetry between outward (export) and inward (import) market access, important implications both for the characterization of equilibrium and for patterns of trade and income. Waugh (2010) shows that asymmetric trade costs might be an important part of the explanation for real per capita income differences between countries. In contrast, unlike in our paper, Brancaccio *et al.* (2020) emphasize that trade costs are shaped by the entire global network, being influenced by factors like shipping prices, distance, and the volume of trade, and therefore substantial differences of trade costs in trade data are generated by such endogeneity. Closer to our paper is a literature on gravity models and general equilibrium under asymmetric trade costs with Bergstrand *et al.* (2013) being a notable contribution. What we contribute to this literature is a new, general theoretical framework to show how asymmetry matters for trade and, in particular, how the empirical evidence of asymmetric trade costs contributes in a quantitatively important way to understanding the location and sizes of Greek cities, and thus to European economic history, geography and urbanization.

Finally, before we get into the details of modeling, it is appropriate to note that our approach presumes that the founders and inhabitants of the Hellenic poleis were aware of

[11] The historical evidence suggests that, not unlike contemporary monetary unions, Delian League members were more likely to interact with other members than with non-members, which is consistent with the "top-down" nature of the relationship with the hegemon. Other looser and lesser known associations of poleis in the form of regional federations, known as *Koinon* and of which several existed, were more likely to interact extensively in a horizontal fashion with other poleis that were non-members; Mackil (2013).

the geographical expanse of the Hellenic system of cities, which spread from Colchis in modern Georgia to Gibraltar. But, is it appropriate to adopt such an assumption about information? That members of the elites, educated male citizens of Hellenic poleis, did have such a perception of geography is evidenced by the historical record. Herodotus was very well acquainted with far away places from Babylon and Egypt to the Atlantic coast, and from the Mediterranean to the Land of the Hyperboreians (Finland?). Also, a passage from Plato's *Phaedo*, 109a-109b, is quite suggestive:[12]

" 'And rightly,' said Simmias. 'I believe that the earth is very large and that we who dwell between the pillars of Hercules and the river Phasis live in a small part of it about the sea, like ants or frogs about a pond, and that many other people live in many other such regions. For I believe there are in all directions on the earth many hollows of very various forms and sizes, into which the water and mist and air have run together ...' "

The Hellenic system of cities was, in a way, a sort of "natural experiment,"[13] whereby poleis located primarily in coastal regions with similar geography and culture developed a remarkable, politically decentralized (in spite of their similarities and alliances among them) and quite advanced economic system, geographically dispersed on narrow coastal plains and based on trade, with other poleis and with the hinterland, exploiting division of labor and utilizing numerous technological but also institutional innovations [Castoriadis (1991)]. The similarity among the locations of the poleis contrasts sharply with the differences in how long it would take them to reach each other, since winds and their seasonal variation substantially affected sailing times. How important *inter alia* were these factors for the remarkable development of the ancient Hellenic poleis? These are some of the questions that we seek to illuminate.

3 The Economy of the Ancient Hellenic Poleis

The ancient Greek poleis were historical examples of city-states, a form of political organization found in many periods and regions of world history [Woolf (2020), pp. 172–173]. A polis, however, was not simply a city in the modern sense. The term encompassed the body of citizens, the territory they occupied, and the political community through which they governed themselves.[14] Hansen and Nielsen's (2004) *Inventory* identifies 1,037 poleis that

[12]We owe this reference to Ober (2015), Ch. 2. The significance of cultural closeness of the ancient Hellenic cities for the fact that the Hellenic world remained decentralized is first posed in ibid., Ch. 2, and is revisited in Ch. 3 in connection with Aristotle's view of individuals as "political animals."

[13]c.f. Acemoglu *et al.* (2016).

[14]In the ancient Greek conception, most clearly formulated by Aristotle in *Politics*, a polis was "a community (*koinonia*) of citizens (*politai*) with regard to the constitution (*politeia*)." Although this definition

are known from archaeological and historical evidence to have existed during the Archaic and Classical periods, from 650 BCE to 323 BCE. These poleis were self-governing political units and therefore provide a natural setting in which to study trade among sovereign economies.

Poleis traded extensively with one another, with non-Greek-speaking city-states, and with more distant regions. Historical accounts describe the wide geographic reach of the markets supplying larger poleis and document imports of luxury, strategic, and basic commodities [Kallet and Kroll (2020), pp. 28–29]. Archaeological and historical evidence also records substantial trade in agricultural products—especially cereals, olive oil, and wine—as well as in textiles, tools, and other manufactured goods. Because not all poleis issued coinage, transactions and trade balances were settled through a variety of monetary and nonmonetary mechanisms, sometimes including coercive ones.

Most poleis were relatively small, apart from a limited number of major urban centers. Population and settlement-size estimates are available for only a subset of locations and must be interpreted cautiously; see Ober (2015, p. 32, Table 2.1), Hansen and Nielsen (2004), Bresson (2016), Amemiya (2007), and Woolf (2020). These limitations reflect the uneven archaeological record: not all potential sites have been excavated, not all findings have been published, and not all available material has been assembled in a form suitable for systematic analysis. Of the 1,037 poleis identified in the historical record, 143 cannot be assigned a known location, and size information is available for only 635. These gaps determine the available data.[15]

4 Theoretical Modeling

The rudimentary discussion of models of trade that follows in this section fixes ideas, sets notation and motivates our empirical analysis. The relevant literature distinguishes two principal cases: one, the *trade model*, with labor endowments in each location being exogenous and perfectly inelastic; and two, the *economic geography* or *spatial equilibrium model*, with labor in each location being perfectly elastic and a given total population being allocated so as to equalize welfare (real utility) across all locations (Allen and Arkolakis 2014; Allen, Arkolakis and Takahashi 2019, p. 399; Allen and Arkolakis 2025). We appeal to both models. We follow the framework of the latter and after we establish notation we develop a number

initially omits territory, Aristotle later treats the place in which citizens live as an important characteristic of the polis.

[15] Our study started before the Poleis data became available online as an app, Polis-Inventory. Miniscule discrepancies in the counts reflect different decisions in cleaning the original data and are inconsequential.

of propositions that elaborate the properties of the model and help anchor our empirical investigation. In sum, these are asymmetric and seasonally varying trading costs and their implications for the role of market access in determining location and size of trading poleis (or sites). We also explore the consequences of departure from iceberg trading costs.

Labor L_i is the only factor of production in city, polis or site i, the terms being used interchangeably at this point. A single tradeable output is produced in each site i by a large number of producers with identical linear in L_i production functions:

$$Q_i = A_i L_i, \; i = 1, \ldots, J. \tag{1}$$

TFP A_i could be specified in more detail, if necessary, in order to express geographical features, soil conditions, etc. We work with a representative individual at each site, whose preferences are stated in the form of a CES function of $c_{ji}, j = 1, \ldots, J$, the quantities of the goods produced in site j that are shipped to and consumed in site i:

$$U_i := u_i \left(\sum_{j=1}^{J} c_{ji}^{1-\frac{1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where u_i denotes site $i-$specific amenities, and the constant elasticity of substitution $\sigma > 1$ is the same across all differentiated goods that are produced in the system of poleis, $j = 1, \ldots, J$. The indirect utility corresponding to U_i is given by:

$$\Omega_i = u_i \frac{w_i}{P_i}, \tag{3}$$

where w_i is the competitive equilibrium nominal wage rate, $w_i = p_i A_i$, p_i is the output price of good i at site i, also known as the "factory-gate" price, and P_i the "ideal", i.e. Dixit-Stiglitz price index, which will be defined shortly. For a given set of sites occupied by poleis, we invoke the Armington assumption [Armington (1969)] that goods are differentiated by origin, and therefore the range of goods is determined by the number of locations and thus given.

Trading is subject to asymmetric bilateral iceberg trade costs, where $\tau_{ij} > 1$, denotes the units of the good produced at i that must be shipped in order for one unit to arrive at site j. Bilateral prices, that is, the prices that consumers at site j pay for a unit of good i, are given by $p_{ij} = p_i \tau_{ij}$. By expressing local prices in terms of the respective wage rates, the bilateral prices are given by,

$$p_{ij} = \frac{w_i}{A_i} \tau_{ij}. \tag{4}$$

Consequently, the price index P_i which is defined in terms of the bilateral prices p_{ji} for the goods shipped from all other sites $j \in \mathcal{S}$ and consumed at site i is given by:

$$P_i \equiv \left[\sum_k \left(\tau_{ki} \frac{w_k}{A_k} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \quad (5)$$

Using (5) in (3) allows us to express indirect utility Ω_i, which is homogeneous of degree zero in prices and wages, in terms of relative wages only, $w_j/w_i, j \neq i$, that is in real terms.

If polis populations are given, the intercity trade equilibrium determines relative wages uniquely. If individuals are free to relocate, which would make labor supply infinitely elastic at every site, then an alternative conceptualization of an intercity trade equilibrium as a *spatial equilibrium* follows. That is, populations move so as to equalize utilities across all sites, and as a result a given total population L allocates itself over all sites. This is the case that modern spatial economics refers to as the Rosen-Roback assumption, or alternatively the case of economic geography. In that case, trade equilibrium and spatial equilibrium together determine equilibrium prices and populations. We follow Allen and Arkolakis (2025), Section 4.4., and assume that amenities and productivities are constant parameters, that is there are no agglomeration nor congestion externalities.

The assumption of free mobility underlying spatial equilibrium is best understood as a long-run approximation. Although movement across poleis was subject to institutional and social frictions, [16] these frictions did not prevent population from adjusting across locations over extended periods. Persistent differences in economic opportunities and living conditions could therefore affect the long-run distribution of population, even if adjustment was neither immediate nor costless.[17]

4.1 Propositions and Empirical Implications

Proposition 1 (Population and eigenvector centrality) *Equilibrium relative population is a loglinear function of right and left Perron–Frobenius eigenvector centralities.*

As in other spatial equilibrium models, the model determines relative rather than absolute

[16] Mobility conditions varied substantially across poleis. Athens, for example, had a large population of metics, free noncitizen residents who had migrated from other poleis or foreign lands, whereas Sparta imposed much stronger restrictions on outsiders.

[17] Similar predictions for the long-run spatial distribution of population may also be obtained from a Malthusian model with endogenous population. Explicitly modeling migration frictions would require a dynamic spatial framework and is beyond the scope of the present paper; see Kleinman, Liu, and Redding (2023).

populations. Fixing total population at $\bar{L}$, equilibrium population can be written as

$$\ln L_i = \ln\left(\frac{\bar{L}}{\sum_j \tilde{x}_j(\tilde{y}_j)^{\frac{\sigma}{\sigma-1}}}\right) + \ln \tilde{x}_i + \frac{\sigma}{\sigma-1}\ln \tilde{y}_i, \tag{6}$$

where $\tilde{x}_i$ and $\tilde{y}_i$ are, respectively, the right and left Perron–Frobenius eigenvectors of

$$\tilde{\mathbf{A}} = [\tilde{A}_{ij}], \qquad \tilde{A}_{ij} = u_j^{\sigma-1}A_i^{\sigma-1}\tau_{ij}^{1-\sigma}. \tag{7}$$

This result follows from Allen and Arkolakis (2025), Section 4.4.

Empirical implementation. We compute the eigenvectors of the bilateral-proximity matrix $\mathbf{T} = [T_{ij}]$, a component of $\tilde{\mathbf{A}}$, where $T_{ij} := \tau_{ij}^{1-\sigma}$, rather than those of $\tilde{\mathbf{A}}$, for two reasons. First, the eigenvectors of $\mathbf{T}$ provide pure measures of connectivity and are straightforward to interpret. Second, there is no obvious way to partition the available local characteristics between proxies for productivity, A_i, and amenities, u_i. We therefore use proximity-based centrality measures and include the observed local characteristics separately, allowing the data to identify their associations with urbanization; see equation (22) and related discussion further below.

Because $\mathbf{T}$ is directed, its right and left eigenvectors need not coincide. We interpret them as outward- and inward-oriented centralities, respectively. We also use Kleinberg's authority and hub centralities because they provide alternative measures designed specifically for directed networks and distinguish between locations connected to influential destinations and locations accessible from influential origins. They are defined by

$$\mathrm{Au}_i = \sum_j T_{ji}\ \mathrm{Hub}_j, \qquad \mathrm{Hub}_i = \sum_j T_{ij}\ \mathrm{Au}_j. \tag{8}$$

Equivalently, authority and hub centralities are the Perron-Frobenius eigenvectors of $\mathbf{T}^\top\mathbf{T}$ and $\mathbf{T}\mathbf{T}^\top$, respectively. We use them as robustness measures of inward- and outward-market access, that is conceptualized with respect to the trading network, rather than as exact structural substitutes for the left and right eigenvectors of $\mathbf{T}$.[18]

Corollary 1 (Population and market access) *Equilibrium population is also a loglinear function of import and export market access.*

[18] Chen and Ioannides (2026), Section 3.5, make similar use of eigenvector centrality measures in studying the diffusion of coinage and script adoption across Hellenic poleis.

Using the standard definitions

$$\mathrm{IMA}_i := P_i^{1-\sigma} = \sum_k \left(\tau_{ki} \frac{w_k}{A_k} \right)^{1-\sigma}, \tag{9}$$

and

$$\mathrm{EMA}_i := \sum_j \frac{\tau_{ij}^{1-\sigma}}{P_j^{1-\sigma}} w_j L_j, \tag{10}$$

equilibrium population can be written as

$$\ln L_i = \sigma \ln \varpi - \sigma \ln u_i + (\sigma - 1) \ln A_i + \ln \mathrm{EMA}_i + \frac{\sigma}{\sigma - 1} \ln \mathrm{IMA}_i, \tag{11}$$

where ϖ denotes spatial-equilibrium utility.

Empirical implementation. We use Corollary 1 to guide a theory-motivated reduced-form empirical specification rather than to estimate the structural equation in (11) directly. Equations (9) and (10) contain endogenous quantities for which we lack direct measures, including wages, price indices, and productivities at all sites. We therefore construct two empirical analogues of import and export market access. First, we define geography-based market-access measures, denoted by IMA_i^G and EMA_i^G, as unweighted sums of bilateral proximity to all other coastal segments. Like the eigenvector centralities discussed above, these measures capture a location's directional position in the trading network using only geography and imputed travel costs. We use these geography-based measures both as reduced-form measures of directional trade potential and as instruments. Second, we define population-weighted market-access measures, denoted by $\mathrm{IMA\_pop}_i$ and $\mathrm{EMA\_pop}_i$, which weight access to other segments by observed settlement population. These variables are closer to conventional empirical market-access measures but are potentially endogenous, so we account for their endogeneity using the instrumental-variable strategy described below.

Proposition 2 (The no-trade restriction) *In the absence of intercity trade, equilibrium population and economic activity depend only on a location's own characteristics.*

Without trade, only the locally produced good is consumed, so that $C_{ii} = A_i$. Real output and income at polis i are equal to $A_i L_i$, while utility per person is equal to $u_i A_i$. These quantities therefore depend only on the local characteristics u_i and A_i and are independent of characteristics elsewhere in the urban system and of bilateral travel costs.

Empirical implication. This result provides an identifiable restriction.[19] In regressions of urbanization on local characteristics, statistically significant effects of characteristics at other locations constitute prima facie evidence that intercity trade contributed to urbanization.

Proposition 3 (Directional centrality) *Equilibrium population is more responsive to inward-oriented centrality than to outward-oriented centrality.*

Equation (6) can be written as

$$\ln L_i = a_0 + a_1 \ln \tilde{x}_i + a_2 \ln \tilde{y}_i, \qquad a_1 = 1, \qquad a_2 = \frac{\sigma}{\sigma - 1} > a_1, \tag{12}$$

where a_0 collects terms that do not vary across locations. Thus, the model assigns a larger coefficient to left, or inward-oriented, centrality than to right, or outward-oriented, centrality. Equation (11) gives the corresponding market-access result: IMA_i enters with a larger coefficient than EMA_i.

Remark 1 (Scale-dependent trading costs) *The relative importance of export market access may be greater when effective trade costs decline with shipment scale. For example, if* $\tau_{ij}^{\text{eff}} = \tau_{ij}(\text{shipments}_{ij})^{-\eta}$, *where* $\eta > 0$, *then bilateral proximity is multiplied by* $(\text{shipments}_{ij})^{(\sigma-1)\eta}$, *strengthening the role of export market access. A similar effect may arise when exporting becomes viable only after shipments exceed a minimum threshold.*

Empirical implication. We estimate specifications that include inward- and outward-oriented centrality separately and test whether inward-oriented centrality has a larger association with urbanization. The corresponding market-access regressions compare the estimated coefficients on IMA_i^G and EMA_i^G as well as IMA_pop_i and EMA_pop_i.

Because our empirical measures are proxies for the structural objects, we do not impose the exact structural coefficients in equation (12). The scale-dependent-cost mechanism in Remark 1 also provides a possible explanation if the estimated coefficient on export market access exceeds that on import market access.

Proposition 4 (Overall centrality and directional asymmetry) *Under a local approximation, equilibrium population can be expressed as a function of overall centrality, the direction of asymmetry, and the magnitude of asymmetry.*

[19] Although climatic conditions may be correlated across broad regions of the study area, they also exhibit considerable local variation, especially in the microclimates affecting navigation; see Beresford (2012). Local climatic conditions therefore do not necessarily proxy for conditions at other sites.

Decompose the structural matrix as

$$\tilde{\mathbf{A}} = \mathbf{S} + \mathbf{B}, \qquad \mathbf{S} = \frac{\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top}{2}, \qquad \mathbf{B} = \frac{\tilde{\mathbf{A}} - \tilde{\mathbf{A}}^\top}{2}, \tag{13}$$

and consider $\tilde{\mathbf{A}}(\varepsilon) = S + \varepsilon B$, where $\varepsilon = 0$ gives the symmetric benchmark and $\varepsilon = 1$ gives the observed matrix. Let v denote the normalized Perron-Frobenius eigenvector of $\mathbf{S}$.

For a sufficiently small $\delta > 0$, let $x_+(\delta)$ and $x_-(\delta)$ denote the consistently normalized right Perron-Frobenius eigenvectors of $\mathbf{S} + \delta\mathbf{B}$ and $\mathbf{S} - \delta\mathbf{B}$, respectively, and define

$$h \approx \frac{x_+(\delta) - x_-(\delta)}{2\delta}, \qquad z_i = \frac{\varepsilon h_i}{v_i}. \tag{14}$$

For sufficiently small asymmetry,

$$\ln L_i \approx a_0 + (a_1 + a_2)\ln v_i + (a_1 - a_2)z_i - \frac{a_1 + a_2}{2} z_i^2 + e_i. \tag{15}$$

Here, $\ln v_i$ captures overall centrality, z_i captures the direction of asymmetry, and z_i^2 captures its magnitude.

Alternatively, define the arithmetic mean and signed relative difference of the observed right and left eigenvectors as

$$m_i = \frac{\tilde{x}_i + \tilde{y}_i}{2}, \qquad d_i = \frac{\tilde{x}_i - \tilde{y}_i}{\tilde{x}_i + \tilde{y}_i}. \tag{16}$$

For $|d_i|$ sufficiently small,

$$\ln L_i \approx a_0 + (a_1 + a_2)\ln m_i + (a_1 - a_2)d_i - \frac{a_1 + a_2}{2} d_i^2 + e_i. \tag{17}$$

Here, $\ln m_i$ captures average centrality, d_i captures the direction of the imbalance, and d_i^2 captures its magnitude. Near the symmetric benchmark, $m_i = v_i + O(\varepsilon^2)$ and $d_i = z_i + O(\varepsilon^3)$, although this local equivalence need not hold closely at $\varepsilon = 1$. [20]

Empirical implementation. The purpose of this decomposition is to quantify the relative explanatory importance of overall centrality and directional asymmetry. We therefore do not

[20] When $\delta = 1$, $x_+(1) = \tilde{x}$ and $x_-(1) = \tilde{y}$, so $[x_+(1) - x_-(1)]/2 = (\tilde{x} - \tilde{y})/2$. At $\varepsilon = 1$, this gives $z_i = h_i/v_i \approx (\tilde{x}_i - \tilde{y}_i)/(2v_i)$, which is close to $d_i = (\tilde{x}_i - \tilde{y}_i)/(\tilde{x}_i + \tilde{y}_i)$ when v_i is close to $(\tilde{x}_i + \tilde{y}_i)/2$. However, $\delta = 1$ gives a secant over the full observed range of asymmetry rather than a local derivative at the symmetric benchmark, for which a small δ is preferable.

estimate equation (17) subject to its structural coefficient restrictions. Instead, we estimate

$$\ln L_i = \beta_0 + \beta_m \ln m_i + \beta_d d_i + \beta_{d^2} d_i^2 + \mathbf{X}_i^\top \gamma + e_i, \tag{18}$$

where $\mathbf{X}_i$ contains the additional control variables. We then use a Shapley decomposition of the unrestricted regression R^2 to quantify the shares attributable to overall centrality, directional asymmetry, and the additional controls. For this purpose, $\ln m_i$ forms the overall-centrality component, while d_i and d_i^2 jointly form the asymmetry component.

Proposition 5 (Seasonal variation in travel costs) *Holding average travel costs fixed, small seasonal variation in travel costs is advantageous because it raises expected bilateral proximity.*

Suppose that travel costs are equal to $\tau_{ij} + \varepsilon_{ij}$, where $\mathcal{E}[\varepsilon_{ij}] = 0$ and $\mathcal{E}[\varepsilon_{ij}^2] = \varsigma_{ij}^2$. For sufficiently small seasonal variation,

$$\mathcal{E}\left[(\tau_{ij} + \varepsilon_{ij})^{1-\sigma}\right] \approx \tau_{ij}^{1-\sigma}\left[1 + \frac{\sigma(\sigma-1)\varsigma_{ij}^2}{2\tau_{ij}^2}\right] > \tau_{ij}^{1-\sigma}. \tag{19}$$

The result follows from the convexity of $\tau^{1-\sigma}$ when $\sigma > 1$. Intuitively, when trade can be shifted across seasons, poleis can concentrate shipments in relatively low-cost seasons and postpone trade during relatively high-cost seasons.

Empirical implication. The proposition predicts that, conditional on average travel costs, locations with greater seasonal variation in trading conditions may be more favorable for urban settlement. We test this implication using alternative measures of seasonal variation and examine whether they are positively associated with our measures of urbanization.

4.2 Further Theoretical Implications

The model also yields several additional implications that are not directly testable with our data but may provide useful predictions for future research with richer data on bilateral trade costs, trade flows, and welfare.

Proposition 6 (Directional link improvements) *The marginal welfare value of improving a directional link depends on the product of the relevant left and right Perron–Frobenius weights.*

Let q denote a parameter that perturbs $\tilde{\mathbf{A}}(q)$. By the eigenvalue perturbation theorem,

$$\bar{\Omega}'(q) = \frac{\tilde{\mathbf{y}}(q)^\top \tilde{\mathbf{A}}'(q)\tilde{\mathbf{x}}(q)}{\tilde{\mathbf{y}}(q)^\top \tilde{\mathbf{x}}(q)}, \qquad \frac{\partial \bar{\Omega}}{\partial \tilde{A}_{ij}} = \frac{\tilde{y}_i \tilde{x}_j}{\sum_{k=1}^{J} \tilde{y}_k \tilde{x}_k}. \tag{20}$$

Because $\bar{\Omega}$ represents the transformed equilibrium-utility level, improving the link from i to j need not have the same welfare value as improving the reverse link from j to i.[21]

Proposition 7 (Uniform and localized reductions in trade costs) *A uniform proportional reduction in all bilateral trading costs raises equilibrium utility but leaves relative populations unchanged, whereas a localized reduction may affect both welfare and the spatial distribution of population.*

If all bilateral trade costs are multiplied by the same factor κ, where $0 < \kappa < 1$, then

$$\tilde{\mathbf{A}}' = \kappa^{1-\sigma} \tilde{\mathbf{A}}. \tag{21}$$

The Perron–Frobenius eigenvalue therefore rises, while the right and left eigenvectors remain unchanged. Equilibrium utility consequently increases, but equation (6) implies that relative populations do not change. By contrast, a reduction in only a subset of bilateral trade costs raises equilibrium utility and generally changes the eigenvectors, allowing population to be redistributed across locations.

Remark 2 (Piracy) *Piracy may be modeled as an increase in trading costs. If its effect is uniform across routes, equilibrium utility falls while relative populations remain unchanged. If piracy affects only a subset of routes, equilibrium utility also falls and the spatial distribution of population may change.*

An extreme form of piracy would be removal of a polis from the system. This results mechanically, due to the Perron-Frobenius theorem for the monotonicity for nonnegative matrices, in decrease of equilibrium utility. However, a proper assessment should rely on normalizing the utility function, defined by (3), by dividing the summation by J. In that a case, equilibrium utility decreases only if the fall in the Perron-Frobenius eigenvalue is large enough to offset the gain due to normalization.[22]

[21] We acknowledge a referee who urged us to emphasize this implication.

[22] $\frac{\varpi_{J-1}}{\varpi_J} > \left(\frac{J-1}{J}\right)^{\frac{\sigma}{\sigma-1}}$.

4.3 Some Illustrative Examples

The preceding results characterize how overall network position and directional asymmetry shape equilibrium population in a general system of locations. The following stylized examples illustrate these mechanisms in some simple geographies and relate them more directly to settlement patterns around the Mediterranean.

We first consider a symmetric benchmark to illustrate how geography and trade costs can generate dispersion, clustering, or agglomeration in spatial equilibrium. Suppose, for example, that the J sites are located along a circle and that shipping occurs along its periphery. Let iceberg trade costs take the form $\tau_{ij} = \tau^{d(i,j)}$, where $d(i,j) = |i - j| (mod\ J)$. Under the assumptions that all agents consume all goods produced in the economy and that all sites have homogeneous exogenous productivities and amenities, there are three possible location patterns: uniform dispersion, clustered dispersion and full agglomeration. Clustered dispersion emerges when the uniform location is unstable.[23]

We may extend this model by introducing a modicum of asymmetry roughly corresponding to the actual geographic setting of our study. Let sites in the upper half of a circle trade with other sites in the upper half under iceberg costs that depend on distance as above, and with sites in the lower half by means iceberg costs defined as $\tau_{ij} = \tau_i \tau_j$. A closed-form solution for spatial equilibrium may be obtained under the assumption of within-half circle homogeneity. It is characterized by relative populations tilting toward the half circle with stronger internal access and/or stronger links to the other half. These features are accentuated if a similar setting in the form of an ellipse is assumed. Both results roughly correspond to the pattern of settlements around the Mediterranean.

5 Empirical Setting and Data

Our empirical analysis studies how local geography and a location's position in the intercity trading network shaped the spatial distribution of population and urban settlement. We use the observed locations and sizes of archaic and classical poleis as measures of settlement and population, consistent with the population outcome in the spatial-equilibrium model. Because bilateral trade flows are not observed, we characterize each location's position in

[23]See Allen and Arkolakis (2018) for the location problem set in continuous space. Although in their particular case, the space assumed is an interval on the real line, it is straightforward to see that Equation (29), *ibid.*, p. 455, applies and may be solved under the boundary condition $L_i(z) = L_i(2\pi R + z)$, where R denotes the radius of the circle and z locations along it.

the trading system using network-centrality measures discussed and defined in Proposition 1 and import and export market access constructed from imputed bilateral travel costs. We combine these measures with local geographic, climatic, political, and institutional characteristics.

Our maintained historical premise is that poleis throughout the broadly defined Hellenic world traded with one another. We focus on trade among the poleis in our sample and do not explicitly model their interactions with surrounding hinterlands or with cities outside this system. We nevertheless adopt a broad geographic definition of the Hellenic world, including sites not only in the Aegean and present-day Greece but also along the wider Mediterranean and Black Sea coasts. Some of these cities were not Greek in origin but became Hellenized through migration, settlement, or conquest.

5.1 Data Overview

We define a number of proxies for urban settlements in reference to each coastal segment s , $s \in \mathcal{S}$ where $\mathcal{S}$, $|\mathcal{S}| = 2,737$ denotes the entire coast of the Mediterranean and Black Seas, as defined in terms of segments. Please see Appendices A and B for more detail. We have distances and imputed travel times across all segments and a wealth of local crop suitability and climatic data, which we also refer to as micro-geographic characteristics and all of which are included in our Segments Data. In addition, the segments are linked with a dataset consisting of locations, attributes and characteristics of Hellenic poleis, the ancient Greek cities, to be referred to as the Poleis dataset. These data originate in the Stanford Polis Project, with additional data collected and coded by us.

We discuss next in more detail the two principal datasets that we employ, namely the Poleis and the Coastal Segments Data sets. We note that people did not just appear to populate the poleis. Human settlements of varying urbanicity preceded the poleis, starting with Neolithic settlements and followed by Bronze Age settlements. However, our data pertain to the archaic and classical poleis, the ancient Greek cities that are commonly referred to as city states [Hansen and Nielsen (2004)].

5.2 The Poleis dataset

The Poleis Dataset originates in the *Inventory of Archaic and Classical Poleis* [Hansen and Nielsen (2004), henceforth *Inventory*], a compilation of the entirety of the archaeological and historical data for 1037 archaic and classical Hellenic cities by the *Copenhagen Polis Centre*.[24] From among the available variables of particular interest is Size, defined as the size of the polis territory in the archaeological record (see below for further details), and a number of other variables. We supplement these data with local micro-geographic information on ruggedness, distance to the nearest harbors and the sea, crop suitability and temperature data, which control both for local productivity and amenity characteristics. Since the focus of analysis for this paper is where urban settlements are located, we rely heavily on such information. Thus, while the Poleis Dataset contains 1037 poleis in total, our quantitative analysis will be restricted to the 898 poleis with known locations. Furthermore, as we will detail below, in our empirical setting, we consider the coastal line along the Mediterranean and Black Seas as the potential universe of settlements, and therefore consider coastal settlements, i.e., poleis that are within 10 km to the coast only. This further excludes 327 out of the 898 poleis and leaves us with 571 poleis for the final analyses.

The paper emphasizes economic interactions among archaic and classical Hellenic poleis documented and curated by the *Inventory*. This focus does not imply that poleis interacted only with one another. Like other cities of their era, they also interacted extensively with surrounding hinterlands and with non-Hellenic cities, as Woolf (2020) discusses. Our empirical analysis abstracts from these additional interactions and focuses on connectivity among poleis located near the Mediterranean and Black Sea coastlines. At the same time, the *Inventory* records each polis' degree of Hellenicity, allowing us to describe the composition of the sample. Among the original 1037 poleis, 820 are classified as purely Hellenic, 85 as mixed, and 44 as mostly non-Hellenic; among the 571 poleis that we use, 478 are Hellenic, 48 mixed, 16 mostly non-

5.3 The Coastal Segments dataset

The second set of data collection and imputations involves 2,737 coastal segments that cover the entire coastline of the Mediterranean and Black Seas. We define segment midpoints by

[24] Most of the data were digitized by Josiah Ober and his team of Stanford and Oxford scholars. [http://polis.stanford.edu/]. See Also the Polis-Inventory app.

We ignore, for the purpose of the present inquiry, an additional measure of polis size, that is fame, a continuous-valued variable that records the number of mentions of poleis in the *Inventory*, measured in one-eighth of columns. Not surprisingly, size and fame are quite highly correlated, and fame may be thought of as reflecting the total impact of the entire urban system on a polis. However, fame is a different outcome from size, as it reflects a polis' impact through a combination of the historical and archaeological records and thus likely to reflect a multitude of other factors that pertain to scholarship traditions and the like. Therefore, its endogeneity reflects forces in addition to those of size and we therefore are not using the variable fame in this investigation.

using a map of the coastline of the Mediterranean and placing a point at a regular interval of 15km along the line. The Coastal Segments dataset can be seen as a regular grid of points along the coastline in the spirit of Michaels and Rauch (2019) and Bakker *et al.* (2021). Placing points on the coastline ensures that we do not need to define inland transportation costs when computing pairwise transport costs between segment points. Our goal is to create an evenly spaced grid of points that are a representative sample of the coastline of the area we study. Our grid is indeed evenly spaced if the coastline is perceived as a line. However, in terms of area, the density of grid points in a given area of the map can differ because the coastline is straight in some areas and curved in others.

Studies that create regular grids inland and select by distance to coast (such as Bakker *et al.* 2021) also have uneven distribution of points for the same reason. It is not conceptually clear to us whether a grid that is regular in terms of area should be preferable to one that is regular along the coastline. Econometrically, the difference is one of weighting: An area with more gridpoints is given a higher weight in regressions. These segments are fully identified in terms of geographical coordinates and linked to similar micro-geographic data as those of the poleis dataset discussed above. Maritime travel times across all segments are also imputed and the details of these imputations are given in the Appendix, section A. Figure 1 compares imputed with some historical travel times.

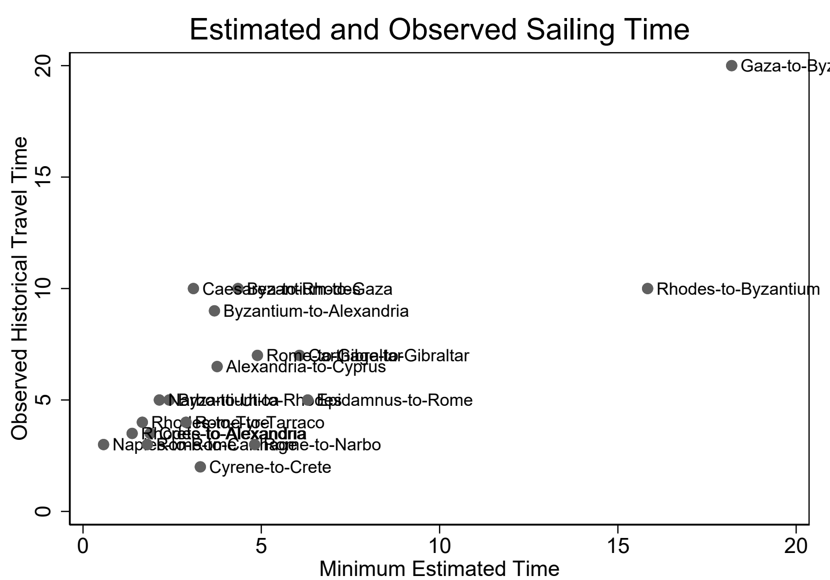
5.4 Imputation of Travel Times

Wind is the central environmental factor in determining performance of ancient vessels. But can contemporary data proxy for ancient conditions? Murray (1987) compares modern wind data to ancient references to wind conditions and suggests that the similarity between ancient and current wind patterns is striking in the Mediterranean. Beresford (2012) casts a more critical view. Specifically, he draws attention to proxy evidence from the pollen record which suggests that the period from the late Iron Age all the way to 200 BCE was characterized by lower temperatures than before and after, which must have affected navigational conditions in the Mediterranean at those times. Those data are not spatially detailed (unlike for other areas of the world), and therefore we cannot do more than caution the reader. However, Beresford and many others agree that the winds much more than currents and the tide determined navigation in the Mediterranean. As Leidwanger (2013) puts it, “a number of wind patterns affect seafaring capabilities, but prevailing seasonal winds were probably the crucial factor; they were both the strongest and the most reliable through the sailing season, and appear most frequently in seafaring accounts.” Therefore, accounting for the seasonal variability is particularly relevant for the present study. Unlike the facts used by Bonnier *et al.* (2020), very little else is known with comparable precision, and there is generally dearth of detailed paleoclimatic data for this region of the world. Our imputations rely on GIS-based methods developed by Alberti (2018). Appendix, section A provides more details.

Imputation of maritime as well as overland travel times has attracted attention by several inquiries. One is Orbis, https://orbis.stanford.edu/, which emphasizes the transportation network of the Roman Empire but falls after the range of coverage of our study. Another inquiry, discussed in the Appendix, is Gal *et al.* (2022), who emphasize the central importance of accounting for seasonality and currents, and also rely on GIS-based methods. Both those sources do not provide the detail that is central to our approach, namely using all coastal segments of the Mediterranean and Black Seas. We report next two data quality checks and leave further details for the Appendix.

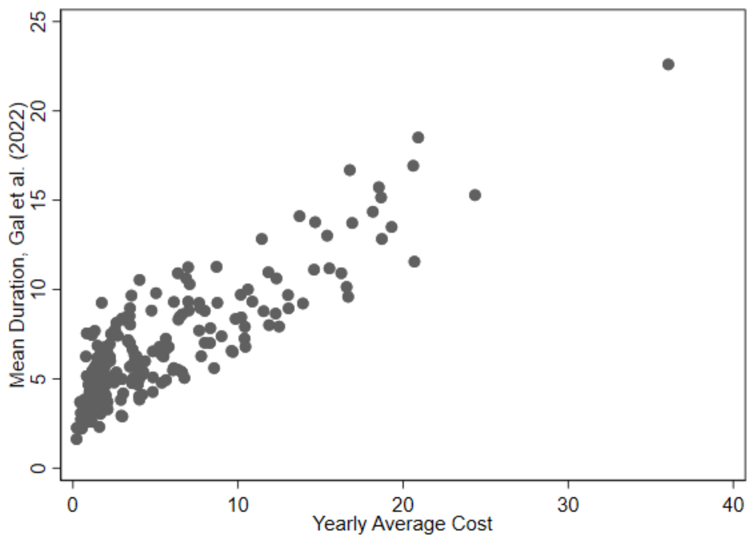
We first check our data quality by comparing our computed sailing times with some selected routes for which there exist records of observed historical sailing times. The correlation between observed actual historical sailing times for a few select routes listed with the calculated sailing times using the above network is shown in Figure 1. Visually they are very consistent. A regression of the minimum (among all seasons) against historical observed minimum voyage times without a constant has a R^2 of 0.82.

Figure 1: Travel Times: Historical Observed vs. Estimated



We also conduct cross-checking with Gal *et al.* (2022), which simulates the sailing times for 224 sailing routes around the Mediterranean. The correlation between our yearly average travel time and their mean duration across all the months is 0.87. The magnitudes are close too, see Figure 2.

Figure 2: Travel Times: Gal *et al.* (2022) vs. Own Imputations



5.5 Measures of Urbanization

The definition of a city, already discussed in the Introduction, is an important one and deserves returning to.[25] The classics literature, and especially Aristotle, see the ancient polis in terms of territory, which includes the physical aspect of concentrated population settlements, of respective populations, and of their political and social organization. The meaning of polis did not change significantly over the archaic and classical period (which the latter by convention ends at 423 BCE). In view of the extensive effort made by Hansen and Nielsen (2004), p. 39–46, 70–73, and their documentation of the classics literature to arrive at a quantitative measure for the concept of the polis, we defer to them and adopt their measure for the purpose of this study.

We measure urban settlement, that is, urbanization, by means of three alternative variables that we define using both the Poleis and Coastal Segments datasets. The first, an extensive margin measure, is the count of poleis settled at each coastal segment, $s \in \mathcal{S}$.

[25]The word polis has a multitude of meanings in the classics literature: (a) stronghold or citadel, (b) nucleated settlement, (c) country or territory, and (d) political community. The modern term city-state adheres to the importance of political institutions, a word that originates in the term polis.

Specifically, we match each polis to any coastal segment that lies within 10km from it. This means that a polis can sometimes be matched to multiple segments. The 571 poleis give us 710 settled segments (thus defined), out of the 2737 segments in total.

In addition to this extensive margin measure, we also use two intensive margin measures. One is the sum of sizes, as reported in the *Inventory*, for all poleis that are within 10 km to the midpoint of a segment. Polis size, as available in the *Inventory*, reflects the broader meaning of a city in ancient Greece, of the respective city-state (a modern term), as discussed above. It is defined in *ibid.* as a categorical variable in reference to following ranges of polis territory: $[0-25), [25-100), [100-200), [200-500), \geq 500$ km^2 [*ibid.* p. 7], recorded as values in $\{1, 2, 3, 4, 5\}$, respectively. It is available for 442 among 569 poleis [*ibid.* p. 71], as discussed in section 5.2. For the 127 poleis without known sizes (records), we assume they all have the smallest size. Arguably, larger poleis are more likely to be linked to written records and to have attracted archaeological attention. Hansen and Nielsen rely on inference that suffices to classify poleis in the above sparse categories while heeding the imprecision of the data.[26] Our treatment assumes that the variation of urban density within the polis territory is minimal, as no data are available in any consistent detail for all of the poleis in our data. While many individual sites have been excavated, the quality of finds in the archaeological record is not sufficiently detailed nor consistent to allow such inference.

The second intensive measure we use is a modification of the size-based measure. That is, for each set of poleis within 10km of a segment, we compute the sum of their populations using estimates by Ober (2015). Specifically, Ober (2015) by relying on data from the *Inventory* and other sources designates larger poleis to size categories 6 and 7.[27] Ober suggests the following mapping from each polis size to population: Polis size 1, population 1,000; size 2, population 3,500; size 3, population 7,000; size 4, population 17,000; size 5, population 35,000; size 6, population 65,000; size 7, population 150,000 [*ibid.*, Table 2.1, p. 32]. Both intensive measures give us a crucial advantage. Aggregating populations effectively over a segment territory circumvents the ambiguity of polis size and of heterogeneity in the density of settlements within the area defined as polis territory.

Given that intercity travel times vary seasonally, it is appropriate to interpret size as the outcome of the long-run average size of the respective city state. With additions and

[26]After inspection of the data, we corrected some inconsistencies, especially regarding GPS coordinates, and augmented the data for some poleis. That investigation revealed that additional data for poleis with missing information might be impossible to obtain at present. Some polis names seem to originate in mere mentions in records at Delphi and other places of worship. All in all, data on size are missing for about one-half of all poleis.

[27]The size 6 poleis are: Argos, Byzantion, Elis, Eretria, Kyrene, Megalopolis, Miletus, Pantikopaion, Rhegium, and Rhodes. The size 7 cities are: Athens, Sparta, and Syracuse.

destructions over the ages and accommodation to the existence of seasonal variations in shipping costs, this is an ambiguous concept. Since shipping costs are available for each of the four seasons, our estimations of outcomes while accounting for seasonal variations enriches the set of factors which the system of cities and the urbanization literature have considered and improves the robustness of our findings. While trade is welfare-enhancing by improving the range of goods available it also exposes the polis economy to seasonal fluctuations in trading costs. Intuitively, polis locations that are particularly advantageous when navigation generally slowed down, as during the winter months, would be favored from the viewpoint of trade. However, as weather conditions vary seasonally, polis sizes reflect "insurance" against normal seasonal weather fluctuations. We return to this issue in section 7.1 below.

We have verified and used the GPS coordinates in the Poleis Data, which originate in the *Inventory*, to merge the Poleis Data with the micro-geographic and spatial information describing the environment surrounding each Hellenic polis. In augmenting the data from the *Inventory* we add, in addition to the geographical descriptions and agricultural suitability of the immediate environment of each polis, distances among sites and imputed travel times; see section 5.4.

Table 1 reports descriptive statistics for the principal variables we use in the empirical analysis. When considering extreme values, it is worth pointing out that the segment point with the highest export market access (EMA) has the coordinates (37.978N, 23.42E), which is about 10km outside of Athens. Thus the city of Athens is close to an extreme value, the point with the best export market access in the entire area we study, clearly a superb location that accords Athens an advantage over export markets and dominance of other coastal points. IMA and EMA have the same mean and standard deviation, which is not surprising as they are both computed as closely related aggregators of the same data points. However, even if they look similar in aggregate, it is worth emphasizing that for individual coastal points the import and export market access can look dramatically different. In Appendix, Figure A6 we show the 10 percent closest points by either sailing direction. Given the strong winds in the Mediterranean, these two sets of points differ widely in the four examples we selected. The correlation of travel times to and from the other coastal points is even negative in those examples, that is, proximity in one sailing direction correlates negatively with proximity in the other direction.

Variable	Obs	Mean	Std. Dev.	Min	Max
polis_count	2,737	0.406	0.808	0	6
ln_pop	2,737	1.137	2.943	0	11.91
IMA_pop	2,737	0.001	0.0012	0.00013	0.0083
EMA_pop	2,737	0.001	0.0012	0.00016	0.0059
Precipitation	2,737	607.4	283.3	58.14	2246.7
Temperature	2,737	16.02	2.75	9.15	21.6
Elevation, mean	2,737	70.36	71.16	0	471.5
Elevation, sd	2,737	99.31	108.04	-3.95	758.2
Malaria	2,737	0.032	0.097	0	2.54
Ruggedness	2,737	1.064	1.103	0	7.198

Table 1: Descriptive statistics of main variables. The unit of observation is a segment mid-point along the coast; *polis_count* is a count of the number of poleis closest to a point, ln *_pop* is the log of $(1 + pop)$, where *pop* is a population estimate based on Ober (2015). *EMA_pop* and *IMA_pop* are population weighted export and import market access measures, respectively, both for the full year. The other variables are control variables, measured for the whole year.

6 Empirical Results

We report in this section econometric results for some of the empirical implications of the propositions developed in the previous sections. Appendix C reports additional results for robustness. They rely on alternative instrumental variables, inspired by the networks literature, and alternative outcome variables for urbanization. We also report results with Poisson regressions that accommodate the discrete nature of the number and sizes of poleis.

Next we report first tests of the basic implication of the model, namely about the role of trade on urbanization, that is Propositions 1 and 2 and Corollary 1. That is, while economic activity at each segment should depend on local productivity and amenity characteristics, local micro-geographic characteristics for short, it should not depend, under autarky, on those characteristics of other segments. Evidence to the contrary is evidence in favor of trade. First, we estimate the importance of trade in the form of the impact on R^2 of the exogenous outward and inward trade potential. We then extend the empirical analysis by means of the broader set of predictions that we discussed earlier, that is, we test the emergence of urbanization as outcome of economic interdependence among population settlements in the geographical area that we focus on. We do so by means of market access regressions that account for the endogeneity of the magnitude and location of urbanization. In doing so we also take advantage of other novel aspects of our data, namely seasonality of trading costs.

Table 1 provides summary statistics of the main variables used in our regressions. Unless we specify otherwise, the number of observations in our regressions is 2737, the number of segment mid-points in our dataset. Variable *polis_count* counts the number of poleis that are nearest in the sense of being within 10km from such a point. These counts vary from 0 to 6. Variable *ln_pop* also sums the number of poleis near segment mid-points, but weights poleis with their population before summation and taking logs. We use polis population estimates from Ober (2015).

6.1 The Importance of Intercity Trade

We use Equation (6) or equivalently Equation (12) to examine the importance of trade. Proposition 1 suggests that log of population is a loglinear function of the left and right eigenvector centralities. We thus examine the importance of trade by looking at the R^2 of the such regressions. We also examine whether the relative magnitudes of the inward and outward access coefficients are consistent with Proposition 3, which predicts that population is more responsive to inward-oriented centrality than to outward-oriented centrality. And we further use Proposition 2 to reject autarky.

Since σ, the elasticity of substitution among varieties, is unknown, we follow much of the trade literature and when we need to we assume $\sigma = 2$. This choice implies that trading costs enter as inverse travel times, which coincides with how trading costs enter as weights in conventional market access measures. Following our above discussion, we assess the importance of trade for urbanization by augmenting Equation (12) by adding $\mathbf{X}_i$, a vector of segment $i-$specific micro-geographic characteristics u_i, A_i without distinguishing between amenity and productivity ones. That is,

$$\ln L_i = a_0 + a_1 \ln \tilde{x}_i + a_2 \ln \tilde{y}_i + \mathbf{X'}_i \boldsymbol{\beta} + \varepsilon_i. \tag{22}$$

It is important to control for these characteristics as suggested by the theory but also as a (partial) remedy for the fact that our computations of $\tilde{x}_i$ and $\tilde{y}_i$ are based only on trading costs across all sites and do not include the amenity and productivity parameters at all sites that enter matrix $\tilde{\mathbf{A}}$, defined in (7). We also follow the discusssion above and estimate a variation of (22) by using the Kleinberg authority and hub centralities instead of the eigenvectors of $\mathbf{T}$.

We estimate versions of Equation (22) by regressing alternative measures of urbanization, such as the log of the count of poleis, log of the sum of sizes and the log of imputed populations, with alternative sets of proxies for trade potential. Table 2, columns 1, 5 and 9 report the R^2 from regressions with only local productivity variables as regressors. Columns 2, 6 and 10 report the R^2 from regressions with the left and right eigenvector centralities being added (which readily follow from our theory); columns 3, 7 and 11 report the R^2 from regressions with the Kleinberg authority and hub centralities, defined in (8), as proxies for trade being added; and columns 4, 8, and 12 report results using geography-based import and export market access, denoted by IMA_i^G and EMA_i^G. These variables are unweighted sums of bilateral proximity to and from all other coastal segments. Throughout these specifications, we set $\sigma = 2$, so that bilateral proximity is equal to inverse travel time. Specifically, the additional regressors are $\ln \text{EMA}_i^G = \ln\left(\sum_{j\neq i} \tau_{ij}^{-1}\right)$ and $\ln \text{IMA}_i^G = \ln\left(\sum_{j\neq i} \tau_{ji}^{-1}\right)$.

Adding the proxies for trade potential improves R^2 for all three urbanization outcomes, with the best fits obtained using the Kleinberg (KC) authority and hub centralities. Adding the eigenvector centralities (EV) increases R^2 from 0.163 to 0.241 for the log count of poleis, from 0.142 to 0.228 for the log sum of polis sizes, and from 0.151 to 0.237 for the log sum of imputed polis populations. The regressions explain a slightly larger share of variation in the count and imputed population outcomes than in summed polis size.

For eigenvector centrality (EV), the ordering of the point estimates is consistent with Proposition 3: inward access has a larger coefficient than outward access across all three urbanization outcomes. The coefficients reported in columns 2, 6, and 10 are positive and statistically significant for both directions, although formal statistical tests do not reject equality of the two coefficients.

The results with the Kleinberg authority and hub centralities (KC) are reported in columns 3, 7, and 11 of Table 2. Compared with the specification with the left and right eigenvector centralities, the regressions with the Kleinberg centralities have larger R^2, suggesting superiority of the Kleinberg centralities.

The results for the geography-based market access measures (MA) are reported in columns 4, 8, and 12 of Table 2. These measures use the same value $\sigma = 2$, so that bilateral proximity is inverse travel time, and aggregate access to and from all other coastal segments without population weights. Their explanatory power is lower than that of both the eigenvector centralities (EV) and the Kleinberg centralities (KC). Nevertheless, all three proxies for trade potential improve the fit, as measured by the estimated R^2, relative to specifications with local characteristics alone. These results provide evidence that intercity trade potential matters for urbanization.

Next we test Proposition 2: in the absence of trade, amenities and productivities at other segments should not matter once we control for the characteristics of the segment in

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	ln(Count of Poleis +1)				ln(Sum of Size +1)				ln(Sum of Pop +1)			
	GEO	+EV	+KC	+MA	GEO	+EV	+KC	+MA	GEO	+EV	+KC	+MA
ln(LEV/AU/IMA (y))		0.227***	0.0473	0.262***		0.362***	0.119	0.435***		2.245***	0.822	2.753***
		(0.0365)	(0.0723)	(0.0534)		(0.0503)	(0.0980)	(0.0725)		(0.296)	(0.585)	(0.423)
ln(REV/HUB/EMA (x))		0.195***	0.371***	0.269***		0.300***	0.524***	0.407***		1.780***	3.169***	2.484***
		(0.0319)	(0.0691)	(0.0411)		(0.0447)	(0.0915)	(0.0586)		(0.273)	(0.563)	(0.357)
R2	0.163	0.241	0.289	0.238	0.142	0.228	0.272	0.226	0.151	0.237	0.285	0.238
N	2737	2737	2737	2737	2737	2737	2737	2737	2737	2737	2737	2737
P-value: equal coeff.		0.417	0.0171	0.908		0.256	0.0264	0.753		0.161	0.0342	0.624

Table 2: Role of Trade: R^2, OLS Regressions

Notes: The table reports OLS coefficients and R^2 for regressions using alternative urbanization outcomes, indicated at the top of the table, and alternative measures of trade potential. All regressions control for local micro-geographic characteristics: precipitation, temperature, elevation (standard deviation and mean), malaria index, ruggedness, and suitability for barley, millet, summer wheat, and winter wheat. Columns 1, 5, and 9 include only these controls. Columns 2, 6, and 10 additionally include the logs of left and right eigenvector centralities, capturing inward and outward access, respectively. Columns 3, 7, and 11 instead include the logs of Kleinberg authority and hub centralities, capturing inward and outward access, respectively. Columns 4, 8, and 12 instead include the logs of geography-based import and export market access, defined as unweighted sums of bilateral proximity from and to all other coastal segments, respectively. Bilateral proximity is measured as travel time to the power of $1-\sigma$, with $\sigma = 2$. Standard errors, reported in parentheses, are clustered using a 1×1 degree grid. The row labeled "p-value: equal coeff." reports p-values for two-sided tests of equality between the inward and outward access coefficients ($H_0 : a_1 = a_2$ in Equation (22)), using the same clustered covariance estimator. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	(1) prec	(2) temp	(3) elev(std)	(4) elev(mean)	(5) malaria	(6) rugg	(7) Crop:B	(8) Crop:M	(9) Crop:SW	(10) Crop:WW
Panel A: L.H.S.: ln(Count of Poleis +1)										
lnMA_x	0.239***	0.294***	0.277***	0.271***	-0.0262	0.282***	0.204***	0.0988***	0.221***	-0.00751
	(0.0376)	(0.0420)	(0.0307)	(0.0299)	(0.0317)	(0.0289)	(0.0381)	(0.0357)	(0.0359)	(0.0335)
Panel B: L.H.S.: ln(Sum of Poleis Sizes +1)										
lnMA_x	0.365***	0.446***	0.417***	0.408***	-0.0400	0.425***	0.318***	0.166***	0.347***	0.0121
	(0.0571)	(0.0631)	(0.0428)	(0.0419)	(0.0482)	(0.0402)	(0.0566)	(0.0543)	(0.0528)	(0.0478)
Panel C: L.H.S.: ln(Sum of Poleis Populations +1)										
lnMA_x	2.259***	2.726***	2.548***	2.495***	-0.272	2.612***	1.946***	1.004***	2.136***	0.119
	(0.354)	(0.384)	(0.259)	(0.253)	(0.298)	(0.242)	(0.354)	(0.346)	(0.334)	(0.298)
N	2737	2737	2737	2737	2737	2737	2737	2737	2737	2737

Table 3: Evidence of Trade: OLS Regressions for Market Access, Local Characteristics of Segments

Notes: This table reports OLS results for regressions measures of urbanization as the outcomes and log of market access to geographical features, defined as the sum of inverse export travel times weighted by the specified micro-geographic characteristic, reported in the top row of the table, that is, precipitation (prec), temperature (temp), elevation-standard deviation (elev(std)), elevation-mean (elev(mean), malaria index, ruggedness (rugg), barley suitability (Crop:B), millet suitability (Crop:M), summer wheat suitability (Crop:SW), and winter wheat suitability (Crop:WW). We suppress the estimated coefficients of the ten local characteristics listed above for i itself in order to save space. Panels A, B, and C employ the same specifications but different measures for urban settlement as the dependent variable, as detailed in the table. Standard errors clustered by 1 degree by 1 degree grid. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

question. We thus estimate the following equation:

$$\ln L_i = c_0 + c_1 \ln\left(\sum_{j\neq i} \frac{x_j}{\tau_{ij}}\right) + \mathbf{X_i'}\boldsymbol{\beta} + \varepsilon_i, \tag{23}$$

where the market access-style variable $\sum_{j\neq i}\frac{x_j}{\tau_{ij}}$ aggregates micro-geographic characteristics of segments $x_j, j \neq i$ by weighting by the respective travel times $\tau_{ij}, d_{ji}\ j \neq i$. If intercity trade is not important for urbanization, c_1 in (23) should be zero. The regression results are reported in Table 3. Columns 1–10 report in turn results with the ten micro-geographic characteristics as x_j, as indicated by the column headings. We can see that out of the 10 variables, 8 have significant estimated c_1's. Only malaria and winter wheat suitabiity do not yield a significant coefficient for the respective aggregate variables.

In sum, both our approaches to assessing the importance of intercity trade for urbanization do confirm that it is indeed important. That is, in regressions with observations on coastal segments and while controlling for the micro-geographic characteristics at each segment we find that including as explanatory variables either several proxies for the potential advantage of trade with all segments and, alternatively, market access style aggregates of the micro-geographic characteristics of all segments improves significantly the R^2's of the respective regressions. In those regressions we do not weight the market access variables by population, which would induce endogeneity. We do handle such endogeneity by instrumental variables methods. Those results follow immediately below.

6.2 Market Access and Urbanization: The Importance of Endogeneity and Asymmetric Travel Times

Corollary 1 and Equation (11) express equilibrium population at each location i in terms of local characteristics and the structural export and import market-access terms, EMA_i and IMA_i. In the empirical analysis, we do not observe all objects entering these structural terms, including wages, price indices, and productivities at all sites. We therefore use commonly used population-weighted proxies for market access, denoted by EMA_pop_i and IMA_pop_i, and estimate their relationship with settlement outcomes defined at the coastal-segment level.

We appeal to the definitions of Equations (10) and (9) and define below two reduced-form variables, $\text{EMA\_}pop_i$ and $\text{IMA\_}pop_i$, that are population-weighted market access measures for export and import market access, respectively, and both measured for the full year. Our definitions of EMA_pop_i and IMA_pop_i in the empirical analysis are motivated as follows. The former expresses the fact that individuals at segment i benefit from the demand for their output by all other segments. The latter expresses the fact that they benefit from a greater variety of goods they consume.[28] We thus proceed with the following definitions:

$$\begin{aligned} \text{EMA\_pop}_i &= \sum_j \frac{\text{pop}_j}{\tau_{ij}}; \\ \text{IMA\_pop}_i &= \sum_j \frac{\text{pop}_j}{\tau_{ji}}. \end{aligned} \tag{24}$$

These are standard population-weighted market access measures, with travel times raised to -1, which correspond to the value of $\sigma = 2$ in the definition of the utility function. A coastal point with a high EMA_pop$_i$ makes it easy to reach other destinations from it. A point with a high IMA_pop$_i$ is easy to be reached by others. Both forms of access can be valuable. Odysseus is more likely to wash up on the shores of the island of Aeaea, the mythological home of goddess-sorceress Circe, if this island has a high IMA_pop$_i$.

We use the universe of coastal segments to test the hypotheses developed in the previous section. We estimate the following adaptation of Equation (11). In particular, we define

$$\ln L_i = a_0 + a_1 \ln \text{\_EMA\_pop}_i + a_2 \ln \text{\_IMA\_pop}_i + \mathbf{X'}_i \boldsymbol{\beta} + \varepsilon_i, \tag{25}$$

where a_0 incorporates the first three terms on the R.H.S. of (11), and $\mathbf{X}_i$ is a vector of site $i-$specific microgeographic characteristics. It is important to control for these for two reasons: one, as a (partial) remedy for the fact that our computations of EMA_pop_i and IMA_pop_i are based only on shipping costs across all sites and do not include the amenity and productivity attributes at each of the sites that enter wages and prices at all sites; second, local amenities u_i and productivity A_i enter directly into the R.H.S. of (11).

In estimating (25) we account for spatial correlation by clustering standard errors at the level of a regular grid at 1×1 degree latitude and longitude. We report our estimates in Table 4. Column 1 reports results with a simple OLS regression using the count of poleis as dependent variable. We think this proxies for the real income in the respective segment, which enters the L.H.S. of (25). In the Appendix, Section C.3, we show that our results hold

[28]When we invoke the Armington model, the number of goods is given by the number of trading sites; Armington (1969). However, in more general models, like those reviewed by Redding and Rossi-Hansberg (2017) and Redding (2022), the number of goods available in the economy varies with the sizes of sites. Consequently, the greater the size of a site from where a given site is importing, the greater the variety of goods it is producing and the lower the associated price index. So, IMA_pop$_i$ depends positively on the sizes of the sites from which i is importing.

when using two plausible alternative outcome variables. Both measures of market access are strongly statistically significant. The coefficient on export market access is more than twice as large in terms of magnitude than the coefficient on import market access. We report a highly similar result in Column 2, where we replace the dependent variable with log population, which proxies for the $\ln L_i$ in (11). The theory predicts that the coefficients a_1, a_2 should be positive while the vector $\mathbf{X'}_i$ of local micro-geographic characteristics is controlled for.

Both market access measures are potentially endogenous in OLS estimations because they depend on the endogenous existence and populations of other segments. A central identification challenge is reverse causality: did high market access cause a major polis such as Athens to grow, or did Athens' size spur development in its neighborhood, thereby raising its measured market access? To address this concern, we instrument EMA_pop_i and IMA_pop_i with geography-based market-access measures, EMA_i^G and IMA_i^G. These instruments count access to all other coastal segments without population weights and therefore depend only on geography and imputed directional travel costs, not on human settlement or economic activity. This approach follows Bakker *et al.* (2021), with the difference that we incorporate wind direction into the computation of both the market access measures and the instruments. The instrument does solve the central identification challenge referred to earlier: reverse causality. To resolve this endogeneity, we instrument man-made market access with a measure of pure geographic advantage. Specifically, our instrument captures market access derived exclusively from coastal-to-coastal segment, independent of human activity. In Appendix section C.2 we show robustness to alternative instrumental variable approaches.

Columns 3 and 4 report the first-stages of these regressions. Unsurprisingly, these instruments are very strong predictors of the main market access variables; both Kleibergen Paap F statistics exceed 124. Columns 5 and 6 show the second-stage estimations, both using the first stage regressions estimated in columns (3) and (4). Coefficients on market access remain quite similar to the OLS estimates, in terms of magnitude, statistical significance and relative importance. Again, export market access is about two times more important than import market access in explaining either the location of poleis, or of population. The difference is statistically significant at 5 percent level in both Columns 5 and 6. To rephrase this finding, it seems that a location is more favorable if it is able to reach others rather if it can be easily reached.

We note that these findings do not agree with the predictions for the relative magnitudes of the respective coefficients according to Equation (11), when we recall that the underlying assumption for $\sigma = 2$. Difference in the exact definitions between Equation (24) and Equations (9) and (10) provides one explanation. Another potential explanation is Remark 1. In a model where effective trade cost decline with shipment scale, the relative importance of export market access may be greater. However, the fact that export and import market access have significantly different estimated coefficients underscores the importance of asymmetry in travel times. We think that the superior fit in Columns 1 versus 2 and 5 versus 6 is due to the fact that populations are imputed following Ober (2015).[29]

	(1) OLS	(2) OLS	(3) First stage	(4) First stage	(5) Second stage	(6) Second stage
L.H.S.	count of poleis	ln_pop	ln_IMA_pop	ln_EMA_pop	count of poleis	ln_pop
ln_IMA_pop	0.215***	0.627***			0.167***	0.514***
	(0.0295)	(0.101)			(0.0450)	(0.130)
ln_EMA_pop	0.439***	1.193***			0.337***	0.996***
	(0.0446)	(0.128)			(0.0442)	(0.136)
Precipitation	-0.000224***	-0.000929***	-0.000353**	-0.000220*	-0.000221***	-0.000929***
	(7.12e-05)	(0.000251)	(0.000161)	(0.000123)	(7.45e-05)	(0.000251)
Temperature	0.0298***	0.0461	-0.00724	-0.0899***	0.0246**	0.0397
	(0.00977)	(0.0292)	(0.0182)	(0.0124)	(0.0114)	(0.0340)
Elevation, mean	0.00193**	0.00776**	0.00183***	0.000991**	0.00212**	0.00813**
	(0.000876)	(0.00317)	(0.000587)	(0.000491)	(0.000925)	(0.00324)
Elevation, sd	-0.00136***	-0.00604***	-0.00132***	-0.000897***	-0.00148***	-0.00629***
	(0.000413)	(0.00179)	(0.000431)	(0.000319)	(0.000431)	(0.00185)
Malaria	0.0532	0.0321	-0.466	-0.387**	-0.0458	-0.179
	(0.0577)	(0.205)	(0.409)	(0.194)	(0.0832)	(0.262)
Ruggedness	0.0101	0.0684	0.123***	0.0593*	0.0406	0.133
	(0.0363)	(0.154)	(0.0393)	(0.0311)	(0.0338)	(0.149)
Crop suitability	0.0120***	0.0237***	0.00379	0.00145	0.0119***	0.0235***
	(0.00312)	(0.00703)	(0.00310)	(0.00229)	(0.00339)	(0.00755)
ln_IMA			1.916***	0.592***		
			(0.114)	(0.0785)		
ln_EMA			0.167	1.535***		
			(0.115)	(0.0796)		
Constant	4.466***	13.39***	-13.35***	-11.94***	3.465***	11.25***
	(0.379)	(1.162)	(0.509)	(0.357)	(0.473)	(1.330)
Observations	2,737	2,737	2,737	2,737	2,737	2,737
First stage F (Kleibergen-Paap)			124.3	124.3		
R-squared	0.319	0.182	0.702	0.700	0.306	0.178

Table 4: OLS and IV Regressions: Urbanization against Import and Export Market Access

Columns 1 and 2 show OLS results. Columns 3 and 4 show first stages that are both used in the second stages shown in Columns 5 and 6. Both instrumented variables are used jointly in 5 and 6. Standard errors are clustered using a 1×1 degree grid, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[29] Although our theoretical motivation emphasizes trade, inter-city interactions in the ancient Greek world took many forms. Conditional on local micro-geographic controls, ease of reaching other locations and ease of being reached may also have affected settlement patterns through channels such as military capacity, piracy exposure, information flows, political integration, and state capacity. We therefore interpret the empirical market-access measures more broadly as capturing directional connectivity across sites, of which trade is one important channel.

6.3 Importance of Asymmetric Travel Times

The inward and outward measures of centrality and market access combine two distinct sources of variation. First, some locations are better connected overall: they have high access to other locations in both directions. Second, some locations are directionally asymmetric: they may be easier to sail from than to sail to, or vice versa. Proposition 4 provides a way to separate these two components. We use this decomposition to quantify how much of the explanatory power of trade potential comes from overall position in the trading network and how much comes from directional asymmetry in travel costs.

Specifically, for each pair of inward and outward trade-potential measures, we construct an average-access component and an asymmetry component, following Equation (18). We then estimate the unrestricted regression and apply a Shapley decomposition to its R^2. The Shapley decomposition attributes explanatory power to groups of regressors by averaging their marginal contributions across possible orderings. We group the average-access term as "overall location" and the directional-difference terms as "asymmetry." We repeat this exercise for eigenvector centrality, Kleinberg centrality, geography-based market access, and settlement-weighted market access.

Table 5 reports the results for the same three outcome measures used in Table 2: the log count of poleis, the log sum of polis sizes, and the log sum of imputed polis populations. Across specifications, asymmetry in travel costs accounts for about 3–4 percent of the regression R^2. Overall location explains a larger share, ranging from about 27 percent for eigenvector centrality to about 70 percent for settlement-weighted market access. Thus, the main role of geography operates through broad differences in network position, but directional asymmetry provides a non-negligible additional contribution. This finding supports the empirical relevance of asymmetric travel times beyond their effect on the overall level of trade potential.

6.4 Market Access Varying by City Size

We aim at a deeper understanding of the interdependence among poleis by accounting specifically for different polis sizes, as reported in the data, in defining market access. This analysis is reported next.

Table 6 shows a version of our main specification where we vary market access by size. Each coefficient in this table comes from a different regression. These regressions follow the OLS specifications in Table 4, except that market access is separated by size of the city that

	Eigenvector Centrality			Kleinberg Centrality		
	Count	Size	Pop	Count	Size	Pop
Overall location	26.90	29.94	30.72	56.07	59.82	60.55
Asymmetry	3.36	3.16	4.07	4.31	3.41	3.73
Total R2	0.20	0.18	0.19	0.29	0.27	0.29
	MA			MA_L		
	Count	Size	Pop	Count	Size	Pop
Overall location	42.62	47.06	48.31	71.13	72.42	70.72
Asymmetry	3.44	2.72	3.57	3.97	3.99	3.78
Total R2	0.24	0.22	0.24	0.39	0.36	0.35

Table 5: Relative Importance of Location and Asymmetry: Shapley Decomposition

Notes: The table reports Shapley decompositions of R^2 for different urban-settlement outcomes and trade-potential measures. All regressions include local micro-geographic characteristics as control variables. The three outcomes are the same as in Table 2: log count of poleis, log sum of polis sizes, and log sum of imputed polis populations. Overall location reports the share of total R^2 explained by the log average of the two directional access measures, and Asymmetry reports the share of total R^2 explained jointly by the signed asymmetry term and its square, as defined in Equation (18). EV refers to right and left eigenvector centralities, KC to Kleinberg hub and authority centralities, and MA to geography-based market access, EMA_i^G and IMA_i^G, computed as unweighted sums of inverse travel times to and from all other coastal segments. MA_L refers to settlement-weighted market access measures analogous to EMA_pop_i and IMA_pop_i, but using the same settlement measure as the dependent variable as weights. Thus, for the count outcome, MA_L is weighted by poleis counts; for the size outcome, by summed polis size; and for the population outcome, by imputed population. Entries for Overall location and Asymmetry are percentages of total R^2.

is traded with. For example, "MA size 1" in the column for EMA only considers export market access to cities of size 1. The same control variables are used but omitted from this table. Coefficients tend to increase with size of the MA studied, implying that it is more valuable to trade with a large polis rather than a small polis. Note also that we do not use logs for the market access variables, because we want to interpret these coefficients linearly and not as elasticities. These findings show that market access to or from larger cities is more valuable than to or from smaller cities. While not a surprising finding, this sensible result strengthens the credibility of our data and model.

	IMA	EMA
MA size 1	0.350***	0.600***
	(0.0335)	(0.0475)
MA size 2	0.305***	0.403***
	(0.0318)	(0.0379)
MA size 3	0.660***	0.727***
	(0.0776)	(0.0798)
MA size 4	0.674***	0.970***
	(0.0976)	(0.124)
MA size 5	0.709***	0.870***
	(0.0914)	(0.122)

Table 6: Variation of Market Access by City Size

Notes: Each coefficient in this table comes from a different regression. These regressions follow the OLS specifications in Table 4, except that market access is separated by size of the city that is traded with. For example, "MA size 1" in the column for EMA only considers export market access to cities of size 1. The same control variables are used but their coefficients are omitted from this table. The number of observations is 2737 in each regression. Standard errors are clustered using a 1 × 1 degree grid; stars denote significance at 1, 5 and 10 percent level.

7 The Importance of Seasonal Travel Times

The previous section examines annual trading costs, which are based on imputed travel times, measured by averaging travel times across seasons. This section turns to the role of seasonal travel conditions. We first examine whether travel costs in some seasons are more strongly associated with settlement patterns than those in others. We then test Proposition 5, which predicts that, holding average travel costs fixed, greater seasonal variation is advantageous.

We report in Table 7, columns 1-4, estimates that are similar to those in Table 4, columns 5 and 6, with the exception that we now use seasonal variation in travel times to test the relationship by season. In these regressions, too, we control for the same local characteristics.

We show second-stage results only, omitting OLS results and first stages. All market access coefficients are again positive and strongly statistically significant. We note that in all seasons, import market access is less important than export market access in explaining population of a segment point. This gap is most pronounced in Winter followed by Fall, where import market access is less important than in the other seasons, and export market access more important. There was less and less trade as the winter approached, and the absolute and relative importance of export against the import market access steadily increased. This underscores the paramount importance of export market access.

	(1)	(2)	(3)	(4)
L.H.S.	ln_pop (Spring)	ln_pop (Summer)	ln_pop (Fall)	ln_pop (Winter)
ln_IMA_pop	0.683***	0.541***	0.410**	0.397***
	(0.119)	(0.145)	(0.159)	(0.119)
ln_EMA_pop	0.769***	1.006***	1.061***	1.336***
	(0.136)	(0.145)	(0.151)	(0.328)
Precipitation	-0.000896***	-0.000651**	-0.000791***	-0.000858***
	(0.000286)	(0.000284)	(0.000261)	(0.000280)
Temperature	-0.0358	0.0299	0.0432	0.00594
	(0.0271)	(0.0456)	(0.0494)	(0.0262)
Elevation, sd	0.00788**	0.00738**	0.00810**	0.00792**
	(0.00320)	(0.00321)	(0.00327)	(0.00346)
Elevation, mean	-0.00724***	-0.00644***	-0.00645***	-0.00669***
	(0.00188)	(0.00189)	(0.00189)	(0.00200)
Malaria	-0.247	-0.189	-0.238	-0.174
	(0.234)	(0.286)	(0.300)	(0.227)
Ruggedness	0.265*	0.196	0.200	0.238
	(0.145)	(0.146)	(0.159)	(0.157)
Crop suitability	0.0226***	0.0258***	0.0234***	0.0196**
	(0.00769)	(0.00719)	(0.00739)	(0.00759)
Constant	-0.978*	-2.251***	-2.348***	-2.490***
	(0.552)	(0.700)	(0.839)	(0.893)
Observations	2,737	2,737	2,737	2,737
R-squared	0.162	0.175	0.161	0.165

Table 7: Seasonal Travel Times

Notes: Second-stage results, similar to column (6) in Table 4, except separated by season. Standard errors are clustered using a 1 × 1 degree grid, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

7.1 Further Aspects of Seasonal Variation of Trading Costs

Unlike the preceding analysis that addresses the importance of seasonal trading costs, always proxied by travel times, while separating the data by season, the present section addresses the consequences of the seasonal variation of travel times for the number and sizes of cities and their population by segment. Such variation implies seasonal variation in prices faced by individuals and firms. The quasi-convexity of indirect utility and profit functions with respect to prices suggests that such variations are welfare improving in partial equilibrium settings. Proposition 5 predicts such an effect; see Equation (19). However, that is not the case in general equilibrium, where such variation implies seasonal variation in the value of endowments.[30] Although theory does not offer specific predictions regarding the impact of seasonal variations in trading costs in general equilibrium, it is interesting to see empirically their impact on urbanization.

We explore the effect of seasonal variation by means of measures of dispersion in travel times across seasons as additional explanatory variables in regressions along the lines of of Equation (22). At a first level of approximation, we use Equation (19) to compute the certainty equivalents of the terms involving the travel times. As we noted earlier, because of the convexity of those terms, the coefficient of variation is predicted to enter with a positive sign.

We work with several definitions of the coefficient of variation for the market access variables. First is:

$$CV(MA)_i = \frac{\sqrt{1/3 * \sum_t (MA(t)_i - \overline{MA_i})^2}}{\overline{MA_i}}, \tag{26}$$

where $MA(t)_i$ is the overall market access (sum of inverse travel times) for coastal segment i during season t, with $t \in T := \{spring, summer, fall, winter\}$, and $\overline{MA_i}$ is the average market access for i defined over the four seasons. We also work with two alternative definitions of the coefficient of variation and modify Equation (26) accordingly.

For our second definition, instead of $\overline{MA_i}$ we use market access computed at the yearly average travel costs:

$$CV(MA')_i = \frac{\sqrt{1/4 * \sum_t (MA(t)_i - MA(year)_i)^2}}{MA(year)_i}.$$

And for our third definition, the coefficient of variation is computed by averaging directly over the realizations of bilateral travel costs rather than using the precomputed market access variables. That is:

$$CV(\tau)_i = \sum_{k \neq i} \frac{\sqrt{1/3 * \sum_s (1/\tau(t)_{ki} - \overline{1/\tau_{ki}})^2}}{\overline{1/\tau_{ki}}}, \tag{27}$$

[30]We thank Herakles Polemarchakis for help in clarifying this argument. In general equilibrium, the dependence of incomes (that is, the valuations of endowments) on endowments must be taken into account in computing derivatives and second derivatives of indirect utility with respect to prices. Therefore, (quasi)-convexity does not hold. If it did, this would lead to a contradiction of the first welfare theorem: average consumption over variable prices would dominate consumption at the average price.

where k ranges over all segments other than i, with $\tau(t)_{ki}$ being the travel time from k and i during season t. Note that $\tau_{ki}^{1-\sigma} = 1/\tau_{ki}$ for $\sigma = 2$. $\overline{1/\tau_{ki}}$ denotes the mean over the four seasons. Last, we also compute a modified version of (27) where instead of $\overline{1/\tau_{ki}}$ we use the travel cost based on yearly average wind data. That is,

$$CV(\tau')_i = \sum_{k \neq i} \frac{\sqrt{1/4 * \sum_t (1/\tau(t)_{ki} - 1/\tau(year)_{ki})^2}}{1/\tau(year)_{ki}}.$$

We highlight the asymmetric nature of our travel costs by computing separately importer and exporter CV, ICV and ECV, respectively, based on whether the MA or bilateral travel cost τ is computed as *to* i, or *from* i, τ_{ji} and τ_{ik}, respectively.

Whereas these CV definitions are based on the universe of segments $\mathcal{S}$ and are therefore exogenous, we also compute their counterparts pertaining to settled segments. We refer to them as settled segments CVs. For example, for definition (26), the respective settlement CV involves $MA(s)_i$ as the sum of inverse travel times to settled segments only, rather than to all segments. Similarly, for definition (27), the respective settlement CV would sum only over settled segments.

Table 8 reports results of regressions along the following equation:

$$\ln L_i = c_0 + c_1 \ln ICV_i + c_2 \ln ECV_i + c_3 \ln IMA^G(year)_i + c_4 \ln EMA^G(year)_i + \mathbf{X'}_i \boldsymbol{\beta} + e_i, \quad (28)$$

where the dependent variable L_i is either the log of the count of poleis, the sum of poleis sizes, or the sum of poleis populations, and $\mathbf{X}_i$ at each segment is the same set of micro-geographic variables that proxy for amenity and productivity characteristics as those used in previous regressions. We also control for $\ln IMA_i^G$ and $\ln EMA_i^G$ to separate out the effect of seasonal variation from those of average market access. We estimate the above equation with OLS when the CVs on the R.H.S. are the exogenous CVs, and with 2SLS when the CVs on the R.H.S. are the settled segments CVs, using the exogenous CVs as instruments.

The regression results reported in Table 8 show very similar results for the exogenous CVs (columns 1 – 4) with those for the settled segments CVs. The seasonal variation in travel times in the form of ECV has a positive and significant effect when it is measured with all but the first definition of the coefficient of variation ($CV(MA)_i$). The variations for exporting, the ECV's, are numerically more important than those for importing, which is consistent with the relative magnitudes of imports and exports that we discuss earlier in section 6.2. Last, the results for the settled segments are quantitatively similar to the respective ones for the universe of segments. As we discuss earlier, the lack of theoretical predictions on the effects of seasonal variations makes our findings quite provocative.

However, we conjectured in Section 5.5 that larger settlement sizes associated with more pronounced seasonal fluctuations may mitigate risk. Specifically, eschewing a formal derivation, we may argue as follows. Seasonal fluctuations affect all settlements in a correlated fashion. Trade improves welfare but also entails risks via the shipping costs, which due to power utility components of the utility function (3) imply that risk tolerance increases with income [Wilson (1982), p. 215]. Thus, larger sizes afford settlements greater insurance against seasonal fluctuations. This is a notable result within the system of cities literature.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Segments CV (OLS)				Settled Segments CV (2SLS)			
	CV(MA)	CV(MA')	CV(τ)	CV(τ')	CV(MA)	CV(MA')	CV(τ)	CV(τ')
Panel A: L.H.S. ln[**Count of Poleis +1**]								
ln(ICV)	0.0249	0.0398	-0.0258	0.0252	0.0286	0.0397	-0.00487	-0.000987
	(0.0220)	(0.0246)	(0.0697)	(0.0554)	(0.0304)	(0.0372)	(0.0501)	(0.0428)
ln(ECV)	0.0344	0.0801***	0.380***	0.317***	0.0674	0.179***	0.299***	0.273***
	(0.0232)	(0.0252)	(0.114)	(0.0638)	(0.0454)	(0.0590)	(0.0918)	(0.0582)
Panel B: L.H.S. ln [**Sum of Poleis Sizes +1**]								
ln(ICV)	0.0337	0.0504	-0.0498	0.0115	0.0383	0.0471	-0.0162	-0.0191
	(0.0329)	(0.0365)	(0.0959)	(0.0817)	(0.0453)	(0.0558)	(0.0700)	(0.0633)
ln(ECV)	0.0488	0.123***	0.558***	0.435***	0.0955	0.270***	0.438***	0.375***
	(0.0323)	(0.0361)	(0.153)	(0.0912)	(0.0634)	(0.0864)	(0.125)	(0.0830)
Panel C: L.H.S. ln [**Sum of Poleis Populations +1**]								
ln(ICV)	0.0979	0.205	-0.442	0.170	0.0935	0.156	-0.179	-0.0736
	(0.201)	(0.223)	(0.566)	(0.503)	(0.278)	(0.334)	(0.417)	(0.391)
ln(ECV)	0.294	0.730***	4.011***	3.182***	0.562	1.571***	3.144***	2.743***
	(0.211)	(0.233)	(0.908)	(0.557)	(0.405)	(0.536)	(0.751)	(0.514)
Geo var	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln IMA, ln EMA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage F					94.16	71.29	330.29	451.54
N	2737	2737	2737	2737	2737	2737	2737	2737

Table 8: Effects of Seasonal Variation: OLS and 2SLS Regressions

Notes: This table reports results with OLS and 2SLS results regressions that have urbanization measures (as indicated at panel title) as outcomes, and logs of the coefficients of variation (CV) for travel times over four seasons for both imports and exports, as well as log of sum of inverse travel times to/from all segments, and local geographical variables (including precipitation, temperature, elevation (standard deviation), elevation (mean), malaria index, ruggedness, barley suitability, millet suitability, summer wheat suitability, and winter wheat suitability; omitted to save space) as the explanatory variables. We use four different measures of the CVs, as defined in section 7.1. The CVs in columns 1 – 4 are computed using travel times to/from all segments around the Mediterranean and the Black seas, while the CVs in columns 5 – 8 are computed using travel cost to/from settled segments only. Columns (1) – (4) report estimates using OLS and columns 5 – 8 using 2SLS with the corresponding CVs to/from all segments as the instruments. The $p-$values come from Wald tests for the equality of coefficients for logs of market access to urban settlement for imports and exports. The first stage F is the Kleibergen-Paap Wald rkF statistic from the first stage in 2SLS. Standard errors clustered by 1 degree by 1 degree grid. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

An interesting refinement would involve the trading costs, which we handle here by means of a certainty equivalent adjustment. A general equilibrium approach would account for the fact that seasonality implies that all relevant costs affecting all poleis fluctuate jointly. Accounting for this feature of the data is possible, but we leave it for future research.

8 Conclusions

This paper examines the importance of trade for the sizes of Hellenic poleis and patterns of urbanization more generally in the ancient Hellenic world. The research pays particular attention to the consequences of asymmetries in trading costs. We first demonstrate the importance of trade by showing that productivities and amenities at all coastal sites around the Mediterranean and Black Seas matter for urban settlement at every site. That is, we show that ancient urbanization at each site is explained not only by local productivity and amenity characteristics but, in addition, significantly so by measures of trade potential that take into consideration presence of urban settlements everywhere in the entire area of study along with productivity and amenity characteristics and travel times. We employ several proxies for urbanization and vectors of productivity and amenity characteristics.

The central theoretical implication of our model is that asymmetric trading costs create distinct import- and export-oriented measures of trade potential, with different implications for settlement size. The strict version of the model implies that settlement outcomes are log linear functions of the eigenvector centralities associated with the spatial-equilibrium system. In the specification closest to the theory, where we use the left and right eigenvector centralities as regressors, the estimates support the model's directional prediction: import market access (or potential) is more strongly associated with settlement outcomes than export market access. This result is not confirmed, however, when alternative measures of market access are used, such as Kleinberg hub and authority centralities and import and export market access, especially when we account for the endogeneity of settlement size and location. Using a variety of instruments, we obtain a robust result, in many instances though not all, that export market access to settlements is more important than import market access to settlements. In addition, we document that seasonal variation in trading costs is beneficial for the emergence of early poleis, consistent with the idea that settlements could take advantage of favorable shipping seasons.

These facts about trade may be true more generally, if asymmetry were to be emphasized. They are definitely intuitively appealing in the context of the system of the ancient Hellenic poleis many of which thrived as export economies. In addition, our research provides new insights into the factors that determine the location of poleis in ancient Hellas. Our key trade-related measures go much beyond the commonly used local geography-based trade-related covariates, thus providing new insights on the fundamental role of geography in shaping the urban structure of ancient Greece. While we do not claim any methodological innovations, we believe that our application in this paper of the concepts involving the eigenvectors of the trading system in studying location of economic activity is the first empirical demonstration of the power of those tools.

Our findings pose some interesting and rather challenging issues in the context of the literature. Specifically, ancient writers like Xenophon and Aristotle, but also classical and mercantilist writers, all of which embodied political economy considerations, regarded robust exports as the foundation of urban prosperity. On the one hand, they praised exports as linked to cities' abilities to earn silver and accumulate other precious commodities and promote economic and political power outward. Dependence on imports, on the other hand, suggested dependence and weakness. Modern trade theory, as employed above and in the absence of political considerations, and evidence, with Redding and Venables (2004) being a good case in point, turned this intuition around. That is, import access plays a larger role in determining real income, welfare, and city size than export access. The logic is straightforward: having cheap and reliable access to a large variety of goods lowers the local price index and directly raises real wages. Export access matters too, but its impact is usually smaller in magnitude because it affects producers' revenue rather than consumers' cost of living. At the same time, the poor quality of the soil and variable climatic conditions in the hinterland of most ancient Greek cities made the production of such staples as grains and such resources as timber unreliable. Ancient Greek cities, and especially the largest ones, depended heavily on reliable grain imports, especially from the black earth area in what is now Ukraine and Russia, but also North Africa. Those supply routes also posed geopolitical challenges; they had to be protected but also generated rents for cities critically located at choke points, like Byzantium's control of the Bosporus. Overall, access to exports was a big advantage. In fact, as we noted earlier, in our data Athens is an outlier in export market access, which is in agreement with the presence of its manufactured goods, like those beautiful vases, everywhere in the Mediterranean. But its hinterland was famously "thinly soiled" ($\lambda\epsilon\pi\tau\acute{o}\gamma\epsilon\omega\varsigma$) and as a result, its welfare and power dramatically depended on long-distance imports.

There exist numerous other important factors that should have affected Hellenic urbanization and deserve attention in future research. Chief among them could be the role of alliances, such as the Delian League and regional federations, but also destructions, takeovers and other consequences of military campaigns and hostilities, more generally. Such inquiries require further delving into ancient history and political economy and arguably would have to rely on additional tools beyond those favored by economists.

9 References

Abulafia, David. *The Great Sea: A Human History of the Mediterranean.* Oxford: Oxford University Press, 2011.

Acemoglu, Daron, James Robinson, and Barry Weingast. 2016. "Review Symposium. Josiah Ober. *The Rise and Fall of Classical Greece* (2015)", *Perspectives on Politics.*14(4):1146.

Alberti, Gianmarco. 2018. "TRANSIT: a GIS Toolbox for Estimating the Duration of Ancient Sail-powered Navigation." *Cartography and Geographic Information Science.* 45(6):510-528.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics.* 129(3):1085-1140.

Allen, Treb, and Costas Arkolakis. 2018. "Modern Spatial Economics: A Primer." 435–472, Ch. 13. In: By, Lili, Yan Ing, and Miaojie Yu. Eds. *World Trade Evolution.* London: Routledge.

Allen, Treb, and Costas Arkolakis. 2025. "Quantitative Regional Economics." NBER Working Paper 33436. January.

Allen, Treb, Costas Arkolakis, and Yuta Takahashi. 2019. "Universal Gravity." *Journal of Political Economy.* 128(2):393-433.

Amemiya, Takeshi. 2014. *Economy and Economics of Ancient Greece.* London: Routledge.

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation." *American Economic Review.* 69(1):106–116.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review.* 93(1):170–192.

Armington, P. S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *International Monetary Fund Staff Papers.* 16:159–178.

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.

Barjamovic, Gojko, Thomas Chaney, Kerem Cosar, and Ali Hortascu. 2019. "Trade, Merchants, and the Lost Cities of the Bronze Age." *Quarterly Journal of Economics.* 134(3):1455-1503.

Beresford, James. 2012. *The Ancient Sailing Season.* Boston: Brill.

Bergstrand, Jeffrey H., Peter Egger, and Mario Larch. 2013. "Gravity Redux: Estimation of Gravity-Equation Coefficients, Elasticities of Substitution, and General Equilibrium Comparative Statics under Asymmetric Bilateral Trade Costs." *Journal of International Economics.* 89(1):110–121.

Bloch, Francis, Matthew O. Jackson and Pietro Tebaldi. 2021. "Centrality Measures in Networks." January. https://arxiv.org/pdf/1608.05845.pdf

Bonnier, Anton, Adam Izdebski, Tymon Sloczynski, Grzegorz Koloch, and Katerina Kouli. 2020. "Landscape Change and Market Integration in Ancient Greece: Evidence from Pollen Data." *Economic Journal.* 130:2596-2618.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou. 2020. "Geography, Transportation, and Endogenous Trade Costs." *Econometrica.* 88(2):657-691.

Bresson, Alain. 2016. *The Making of the Ancient Greek Economy.* Princeton, NJ: Princeton University Press.

Castoriadis, Cornelius. 1991. "The Greek *Polis* and the Creation of Democracy." In. *Philosophy, Politics, Autonomy.* Oxford University Press. 267-289.

Chen, Yuxian, and Yannis M. Ioannides. 2026. “The Diffusion of Alphabetic Writing and Coinage in the Ancient Hellenic World.” Working Paper. Tufts University. Revised. https://tinyurl.com/4eb256t8

Chronopoulos, Dimitris K., Sotiris Kampanelis, Daniel Oto-Peralas, and John O.S. Wilson. 2021. “Ancient Colonialism and the Economic Geography of the Mediterranean.” *Journal of Economic Geography.* 21(5):717-759.

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

Gal, D., H. Saaronin, and D. Cvikel. 2022. “Mappings of Potential Sailing Mobility in the Mediterranean During Antiquity.” *Journal of Archaeological Method and Theory.* https://doi.org/10.1007/s10816-022-09567-5

Hansen, Mogens Herman. 2006. *Polis: An Introduction to the Ancient Greek City-State.* Oxford: Oxford University Press.

Hansen, Mogens Herman, and Thomas Heine Nielsen. 2004. *An Inventory of Archaic and Classical Poleis.* Oxford: Oxford University Press.

Harris, Chauncy D. 1954. “The Market as a Factor in the Localization of Industry in the United States.” *Annals of the Association of American Geographers.* 44:315-348.

Henderson, J. Vernon, Tim Squires, Adam Storeygard, and David N. Weil. 2018. “The Global Distribution of Economic Activity: Nature, History and the Role of Trade.” *The Quarterly Journal of Economics.* 133(1):357–406.

Hicks, John R. 1969. *A Theory of Economic History.* Oxford: Clarendon Press.

Horn, Roger A., and Charles R. Johnson. 2013. *Matrix Analysis*, 2nd ed. Cambridge University Press, Cambridge.

Kallet, Lisa, and John H. Kroll. 2020. *The Athenian Empire.* Cambridge: Cambridge University Press.

Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs. 2004 “A Global Index Representing the Stability of Malaria Transmission”. *The American Journal of Tropical Medicine and Hygiene*: 486-498

Kleinberg, Jon M. 1999. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM.* 46(5):604–632.

Kleinman, Benny, Ernest Liu, and Steven J. Redding. 2023. "Dynamic Spatial General Equilibrium." *Econometrica.* 91(2), 385–424.

Leidwanger, Justin. 2013. "Modeling Distance with Time in Ancient Mediterranean Seafaring: a GIS Application for the Interpretation of Maritime Connectivity." *Journal of Archaeological Science.* 40:3302–3308.

Ma, John. 2024. *Polis.* Princeton, NJ: Princeton University Press.

Mackil, Emily. 2013. *Creating a Common Polity: Religion, Economy, and Politics in the Making of the Greek Koinon.* University of California Press.

Matsuyama, Kiminori. 2017. "Geographical Advantage: Home Market Effect in a Multiregion World." *Research in Economics.* 71:740-758.

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

Murray, William M. 1987. "Do Modern Winds Equal Ancient Winds?" *Mediterranean Historical Review.* 2(2):139-167.

Nunn, Nathan, and Diego Puga. 2012 "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics.* 94(1):20-36.

Ober, Josiah. 2015. *The Rise and Fall of Classical Greece.* Princeton, NJ: Princeton University Press.

Plato. *Phaedo.* 109a-109b. http://www.perseus.tufts.edu/hopper/text?doc=Perseus %3Atext %3A1999.01.0170 %3Atext %3DPhaedo%3Asection%3D109b

Redding, Stephen J. 2022. "Trade and Geography." In: Gopinath, Gita, Elhanan Helpman, and Kenneth Rogoff. Eds. *Handbook of International Trade.* Vol. 5. 147-217. Elsevier: New York.

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

Redding, Stephen J., and and Anthony J. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics.* 62:53–82.

Scheidel, Walter. 2003. "The Greek Demographic Expansion: Models and Comparisons." *The Journal of Hellenic Studies.* 123:120–140

Stachurski, John, and Thomas J. Sargent. 2022. *Economic Networks: Theory and Computation.* QUANTECON BOOK I. https://networks.quantecon.org/

Stanford University Department of Classics. POLIS. General site

https://classics.stanford.edu/projects/polis

Vahrenkamp, Richard. 1976. "Derivatives of the Dominant Root." *Applied Mathematics and Computation.* 2:29–39.

Waugh, Michael E. 2010. "International Trade and Income Differences." *American Economic Review.* 100(5):2093–2124.

Wilson, Robert. 1982. "Risk Measurement of Public Projects." In: Lind Robert C. *et al.* 1982. *Discounting for Time and Risk in Energy Policy.* Resources For the Future, Inc.: Baltimore MD.

Woolf, Greg. 2020. *The Life and Death of Ancient Cities: A Natural History.* Oxford University Press, Oxford.

Zabel, Florian, Birgitta Putzenlechner, and Wolfram Mauser. 2014 "Global agricultural land resources–a high resolution suitability evaluation and its perspectives until 2100 under climate change conditions." *PloS One.* 9(9): e107522.

Appendix

A Sailing Times across Coastal Segments of the Mediterranean and Black Seas

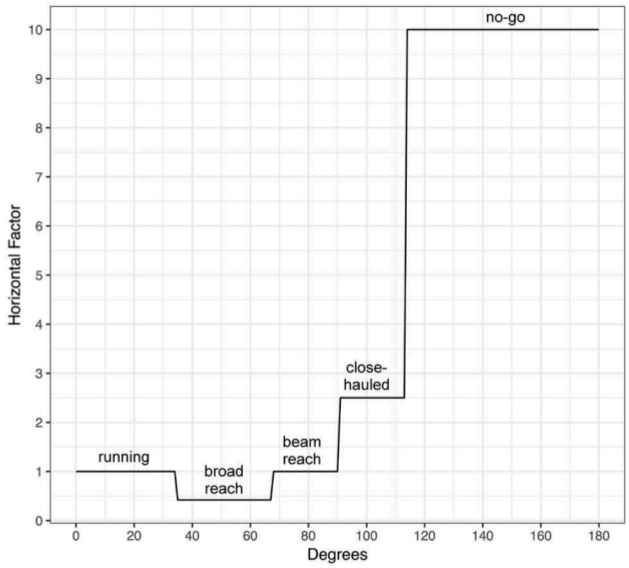
The construction of the dataset of sailing times across all coastal segments was originally carried out by Christopher Barnett, Carolyn Talmadge, Uku-Kaspar Uustalu and the other Tufts Data Lab team. We use the sailing model from Alberti (2018), which computes sailing time given wind speed and wind direction. The directionality of wind is what gives us asymmetric cost. Our wind data for computation from National Oceanic and Atomospheric Administration Global Forecast System (NOAA GFS). Figure A2 shows the wind patterns around the Mediterranean. The data were updated by Dr. Takvor Sukissian and the staff of the Hellenic Center for Marine Research, Greece.

The construction of the sailing cost used the ArcGIS Path Distance tool, and details of it went as follows. First, a cost surface is created in which each cell value is the time to cross a unit distance. According to Alberti (2018), the cost surface is created by scaling the wind speed between the wind minimum and theoretical vessel running speed, then inverting to get time to cross a unit distance. Any land has a "No Data" value, and thus we do not allow inland travel. The cost surface is not directed, but next a horizontal factors raster (wind direction in this case) is combined with it, as a multiplier to the cost surface value. A horizontal factor table is specified which gives the multiplier value corresponding to the angle between the direction of motion and the wind direction.[31] This process allows the cost to be different when traveling in different directions. Figure A1 is the exact value of the horizontal factor used. It shows that the cost could be very asymmetric when sailing from different directions: traveling against the direction of wind takes 10 times more time than traveling in the direction of wind. The ArcGIS Path Distance tool is then applied to the cost surface, with horizontal factors, for each source point, to get surfaces in which each cell is the accumulated least cost path value from the source point. Last, the values at each of the target shoreline points are sampled from output least cost value rasters to get the overall output.

We conduct additional data quality check by cross-checking for a few of our sites which are also available in the Stanford Orbis project (https://orbis.stanford.edu/) for the sailing time

[31]Details: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-the-horizonal-and-vertical-factors-affect-path-distance.htm

Figure A1: Horizontal Factor Table



Source: Figure 3 of Alberti (2018)

in winter. One example is Kranioi and Syrakousai. Figure A3 shows their locations. From Orbis, it takes 5.1 days to travel from Kranioi to Syrakousai, and 4.5 days from Syrakousai to Kranioi. From our data, it takes 6.7 days to travel from Kranioi to Syrakousai, and 1.3 days Syrakousai to Kranioi.

The second example is Chersonesos and Kinolis. Figure A4 shows their locations. From Orbis, it takes 2.2 days to travel from Chersonesos to Kinolis, and 2.7 days from Kinolis to Chersonesos. From our data, it takes 1 days to travel from Chersonesos to Kinolis, and 2.4 days from Kinolis to Chersonesos.

The third example is Euhesperides and Pelusium. Figure A5 shows their locations. From Orbis, it takes 8.7 days to travel from Euhesperides to Pelusium, and 11 days from Pelusium to Euhesperides. From our data, it takes 3.6 days to travel from Euhesperides to Pelusium, and 18 days from Pelusium to Euhesperides.

Figure A6 shows the ten percent closest segment points for a sample of four cities separately for EMA and IMA. The stark differences between the closest points for exports and imports show the strength of travel time asymmetry in this region of the world.

In all these examples, the relative time patterns are consistent between our data and Orbis. The difference in magnitude may come from differences in exact models and wind data

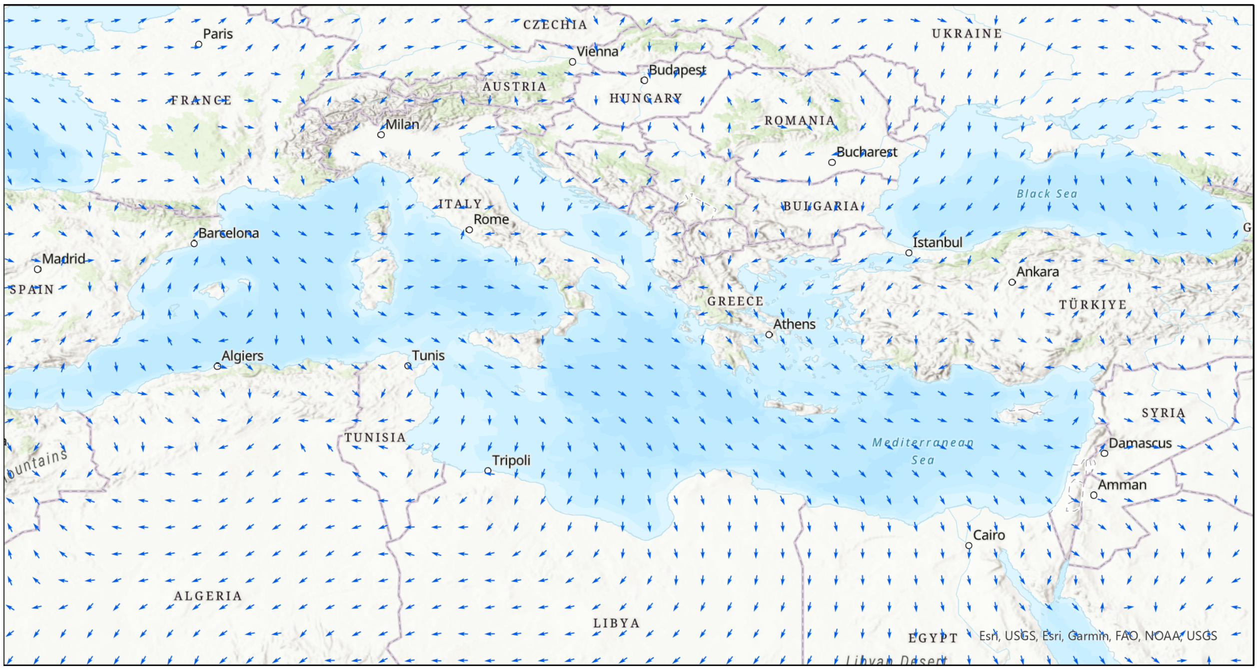


Figure A2: Yearly Average Wind Direction

Note: Arrow indicates the direction that wind goes to.

sources.

B Micro-geographic and Climate Data

The geographic variables are computed as the average value within 5km radius buffer of the center point of the coastal segment. The data sources are the following:

1. The crop suitability data come from Zabel *et al.* (2014). The resolution is 30 arc seconds. The suitability is based on conditions between 1980 and 2019.

2. The ruggedness data come from Nunn and Puga (2012) with resolution of 30 arc seconds.

3. The malaria index is from Kiszewski *et al.* (2004) with resolution of 0.5 degrees.

4. Elevation, precipitation, and temperature data are from WorldClim:

   https://www.worldclim.org/data/worldclim21.html

   The resolution is 30 arc seconds. Both precipitation and temperature are the averages for the years 1970-2000.

5. Data on harbors come from *the Digital Atlas of Roman and Medieval Civilizations*: https://darmc.harvard.edu/data-availability

C Additional Robustness Tests

C.1 The Effects of Export and Import Market Access Using Network Instrumental Variables

We report here results with a method that originates in the network econometrics literature. Following Bramoullé et al. (2009), $\boldsymbol{\mathcal{T}}^2\boldsymbol{x}$ can be used as the instrument for $\boldsymbol{\mathcal{T}}\boldsymbol{y}$, where $\boldsymbol{\mathcal{T}}$ denotes the adjacency (interaction) matrix of the travel network.[32] In our case, $\boldsymbol{\mathcal{T}}$ is defined as $[\frac{1}{\tau_{ij}}]$, and in the place of y_i we have $\widehat{pop}_j$, $\boldsymbol{\mathcal{T}}\boldsymbol{y}$ is just $\sum_j \frac{\widehat{pop}_j}{\tau_{ij}}$. As we have different candidates from among the components of $\mathbf{X}_i$, we use each of them individually, as indicated in the top row of Table A1 to compute $\boldsymbol{\mathcal{T}}\boldsymbol{x}$ and use it as the instrument for $\boldsymbol{\mathcal{T}}\boldsymbol{y}$ or $\sum_j \frac{\widehat{pop}_j}{\tau_{ij}}$. Since the original method is intended for level-level regressions, for this instrumental variable method we estimate a level-level regression and include $\boldsymbol{\mathcal{T}}^2\boldsymbol{x}$ as the instrument and $\boldsymbol{\mathcal{T}}\boldsymbol{x}$ as the control variable, following Bramoullé et al. (2009). Although these network-based instrumental variables are weaker than our other instruments. we do report results because they are based on a different but novel specification.

[32]For social networks with adjacency matrices with entries of 0's and 1's, the square of the adjacency matrix defines the number of paths of length two. For weighted networks like ours, the entries of the square yields the total weight between any two sites through all possible paths of length exactly two. Econometric identification requires that the identity matrix, the adjacency matrix and its square be linearly independent. We thank Yann Bramoullé and Aureo de Paula for related conversations.

	(1) prec	(2) temp	(3) elev(std)	(4) elev(mean)	(5) malaria	(6) rugg	(7) Crop:B	(8) Crop:M	(9) Crop:SW	(10) Crop:WW
Panel A L.H.S.: Count of Poleis										
IMA_L (y)	0.0103	0.0127	0.0663***	0.0512**	0.0105***	0.0484**	0.0123	0.0134***	0.0153**	0.0128***
	(0.00638)	(0.0152)	(0.0257)	(0.0202)	(0.00404)	(0.0202)	(0.00802)	(0.00435)	(0.00657)	(0.00451)
EMA_L (x)	0.0282**	0.0288	-0.0995	-0.0694	0.0312***	-0.0162	0.0260**	0.0307***	0.0234**	0.0116
	(0.0134)	(0.0180)	(0.0663)	(0.0536)	(0.00564)	(0.0316)	(0.0125)	(0.00609)	(0.0106)	(0.0108)
P value($\ln(y) \neq \ln(x)$)	0.283	0.567	0.059	0.081	0.010	0.178	0.457	0.053	0.598	0.925
First stage F	21.77	15.76	7.54	8.43	30.33	14.51	42.28	56.61	42.49	6.05
Panel B L.H.S.: Sum of Poleis Sizes										
IMA_L (y)	0.00114	-0.00459	0.0406*	0.0300	0.0131***	0.0207	0.00846	0.0137***	0.0133**	0.0132***
	(0.00721)	(0.0151)	(0.0211)	(0.0186)	(0.00406)	(0.0199)	(0.00794)	(0.00458)	(0.00597)	(0.00475)
EMA_L (x)	0.0421***	0.0443***	-0.0672	-0.0472	0.0323***	0.00297	0.0366***	0.0325***	0.0322***	0.00803
	(0.0103)	(0.0146)	(0.0480)	(0.0390)	(0.00520)	(0.0245)	(0.0101)	(0.00561)	(0.00826)	(0.0123)
P value($\ln(y) \neq \ln(x)$)	0.007	0.063	0.084	0.123	0.014	0.633	0.091	0.037	0.138	0.715
First stage F	27.71	19.80	10.03	11.40	29.86	14.64	57.67	55.80	59.21	5.50
Panel C L.H.S.: Sum of Polis Population										
IMA_L (y)	-0.0219	-0.0556	-0.000210	-0.00147	0.00731	-0.00533	-0.00923	0.00833	0.00925	0.00293
	(0.0159)	(0.0434)	(0.0323)	(0.0290)	(0.00677)	(0.0292)	(0.0152)	(0.00853)	(0.00778)	(0.0114)
EMA_L (x)	0.0638***	0.0769**	-0.0339	-0.0189	0.0400***	0.00994	0.0571***	0.0365***	0.0371**	-0.00684
	(0.0187)	(0.0304)	(0.0448)	(0.0365)	(0.0112)	(0.0308)	(0.0188)	(0.0130)	(0.0151)	(0.0282)
P value($\ln(y) \neq \ln(x)$)	0.005	0.047	0.592	0.743	0.040	0.754	0.037	0.164	0.182	0.757
First stage F	29.01	20.40	9.86	12.02	28.15	14.65	66.50	51.41	61.42	2.96
Obs	2737	2737	2737	2737	2737	2737	2737	2737	2737	2737

Table A1: Role of Asymmetric Trade Costs and Endogenous Market Access: 2SLS with Network Instrumental Variables

Notes: This table reports 2SLS results for regressions with urbanization measures (as indicated at panel title) as outcomes, and levels of market access weighted by urbanization measures for both imports and exports, as well as local geographical variables (including precipitation, temperature, elevation (standard deviation), elevation (mean), malaria index, ruggedness, barley suitability, millet suitability, summer wheat suitability, and winter wheat suitability; omitted to save space) as the explanatory variables. In each column, $\boldsymbol{G^2x}$ is used as the instrument and $\boldsymbol{Gx}$ is included additionally as a control variable. From Columns (1) – (10), the respective x is: precipitation, temperature, elevation (standard deviation), elevation (mean), malaria index, ruggedness, barley suitability, millet suitability, summer wheat suitability, and winter wheat suitability.
The $p-$values come from the Wald tests for equality of coefficients for logs of market access to urban settlements for imports and exports. The first stage F is the Kleibergen-Paap Wald rkF statistic from the first stage in 2SLS.

Standard errors clustered by 1 degree by 1 degree grid. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C.2 Asymmetric Trade Costs and Endogenous Market Access Revisited: OLS and 2SLS Using Alternative Instruments

Table A2 reports OLS and 2SLS regression results that are an elaboration of those reported in Table 4 in the main text. Relative to that table, we examine robustness by providing regression results with an additional outcome, namely the sum of poleis sizes, and by using a different set of instruments. Column 1 reports OLS results that show that with any of the measures for urbanization exports always have a larger effect than imports. To correct for the potential bias with OLS regressions, Columns 2–7 report results with 2SLS regressions. We use two instrumental variable methods. First, we ignore the variation in pop_j in the terms on the R.H.S. and use $\sum_j \frac{1}{\tau_{ij}}$ as the instrument for $\sum_j \frac{pop_j}{\tau_{ij}}$. Second, we obtain predictions for the pop_j's, $\widehat{pop}_j$'s and use them to construct the instrument. Thus, for the regression reported in column 2, we use the first IV method, namely ignoring the variation in $\widehat{pop}_j$'s and using $\sum_j \frac{1}{\tau_{ij}}$ and $\sum_j \frac{1}{\tau_{ji}}$ as the instruments. The instruments in the regressions reported in columns 3–7 use the second method, namely the variables $\sum_j \frac{\widehat{pop}_j}{\tau_{ij}}$ and $\sum_j \frac{\widehat{pop}_j}{\tau_{ji}}$, where $\widehat{pop}_j$ stands for the predicted value from a Poisson regression of the dependent variable on $\mathbf{X}_i$ only, $\mathbf{X}_i$ and in addition eigenvector centralities (with $\sigma = 1.71$), $\mathbf{X}_i$ and eigenvector centralities (with $\sigma = 2$), $\mathbf{X}_i$ and Kleinberg centralities (with $\sigma = 1.71$), and $\mathbf{X}_i$ and Kleinberg centralities (with $\sigma = 2$), reported in columns 3–7, respectively. The estimated coefficients obtained with 2SLS regressions are typically smaller than those with of their OLS counterparts, perhaps because of endogenous clustering of urban settlements. However, the coefficients are still positive, with those for export being significantly different (according to the $p-$ values) and larger than for import market access, and for all three alternative outcomes. The first-stage $F-$ statistics are all very large. The picture is robust across the results with all outcomes and estimations with different instrumental variables. Finally, comparing the results reported in columns 4 with those in 5, and in 6 with those in 7 suggests that assuming different values for σ makes virtually no difference.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS				2SLS		
		NoL	L:Geo	L:Geo+EV($\sigma = 1.71$)	L:Geo+EV($\sigma = 2$)	L:Geo+KC($\sigma = 1.71$)	L:Geo+KC($\sigma = 2$)
Panel A L.H.S.: Count of Poleis							
ln(IMA_L (y))	0.151***	0.102***	0.116***	0.112***	0.111***	0.114***	0.113***
	(0.0177)	(0.0247)	(0.0227)	(0.0234)	(0.0233)	(0.0224)	(0.0220)
ln(EMA_L (x))	0.214***	0.168***	0.199***	0.201***	0.202***	0.203***	0.205***
	(0.0229)	(0.0227)	(0.0222)	(0.0222)	(0.0224)	(0.0224)	(0.0227)
P value(ln(y) $\neq$ ln(x))	0.061	0.099	0.031	0.021	0.018	0.019	0.016
First stage F		145.58	488.15	563.26	551.62	806.45	862.74
Panel B L.H.S.: Sum of Poleis Sizes							
ln(IMA_L (y))	0.227***	0.177***	0.188***	0.180***	0.179***	0.183***	0.181***
	(0.0256)	(0.0327)	(0.0304)	(0.0311)	(0.0310)	(0.0301)	(0.0297)
ln(EMA_L (x))	0.309***	0.249***	0.288***	0.295***	0.295***	0.297***	0.298***
	(0.0303)	(0.0327)	(0.0304)	(0.0301)	(0.0303)	(0.0301)	(0.0302)
P value(ln(y) $\neq$ ln(x))	0.094	0.192	0.056	0.029	0.027	0.029	0.024
First stage F		162.61	506.59	709.73	687.25	1029.39	1134.85
Panel C L.H.S.: Sum of Poleis Populations							
ln(IMA_L (y))	1.454***	1.256***	1.265***	1.228***	1.219***	1.265***	1.255***
	(0.157)	(0.201)	(0.189)	(0.195)	(0.194)	(0.188)	(0.186)
ln(EMA_L (x))	1.966***	1.652***	1.833***	1.915***	1.918***	1.936***	1.952***
	(0.199)	(0.229)	(0.219)	(0.222)	(0.224)	(0.223)	(0.224)
P value(ln(y) $\neq$ ln(x))	0.098	0.271	0.102	0.053	0.050	0.057	0.049
First stage F		187.91	683.46	705.21	767.18	875.82	876.62
All segments	2737	2737	2737	2737	2737	2737	2737

Table A2: Role of Asymmetric Trade Costs and Endogenous Market Access: OLS and 2SLS Regressions

Notes: This table reports OLS and 2SLS results for regressions with urbanization measures (as indicated at panel title) as outcomes, and logs of market access weighted by urbanization measures for both imports and exports, as well as local geographical variables (including precipitation, temperature, elevation (standard deviation), elevation (mean), malaria index, ruggedness, barley suitability, millet suitability, summer wheat suitability, and winter wheat suitability; omitted to save space) as the explanatory variables. Column 1 is the OLS result. Column 2 is the 2SLS result with sum of inverse travel times to/from all segments (effectively ignoring the numerators $\widehat{pop}_j$) in forming the instrument. Columns 3 – 7 are the 2SLS results with market access weighted by predicted urbanization measures (effectively using predicted rather than actual $pop_j's$ as the numerators) as the instruments. In the regressions reported in columns 3 – 7, the $pop_j's$ are predicted from Poisson regressions that have only geographical variables, geographical variables plus the eigenvector centralities (assuming $\sigma = 1.71$), geographical variables plus the eigenvector centralities (with $\sigma = 2$), geographical variables plus the Kleinberg centralities (with $\sigma = 1.71$), and geographical variables plus the Kleinberg centralities (with $\sigma = 2$), respectively. The $p-$values come from the Wald tests for equality of coefficients for logs of market access to urban settlements for imports and exports. The first stage F is the Kleibergen-Paap Wald rkF statistic from the first stage in 2SLS.

Standard errors clustered by 1 degree by 1 degree grid. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C.3 Alternative dependent variables

In our main specification, our preferred measure for whether a coastal point has urban developments in its neighborhood is a simple count of Poleis near that coastal point. Plausible alternative measures can be conceived. In this robustness section, we show that our main result remains similar if we replace the outcome variable with two plausible alternatives. First, we use a nearest function that assigns each Polis to only its nearest coastal point. We then sum up Poleis that are nearest to each coastal point. This avoids double counting of Poleis from our main measure. Second, we use an indicator function that is equal to one, if there is at least one Polis near a coastal point, and zero, otherwise.

Table A3 shows coefficients of these alternative outcome variables, and compares them with coefficients from Table 4. Coefficient magnitudes change in Panels B and C compared to Panel A. This is expected, as coefficients have different interpretations in each panel. Otherwise, results are very similar across the three panels. Coefficients are positive and significant throughout, and the EMA coefficients are about twice as large as the IMA coefficients.

Table A3: Robustness: Alternative Outcome Measures

	OLS	IV
Panel A: Coefficients reproduced from Table 4		
	count of poleis	count of poleis
ln_IMA_pop	0.215***	0.167***
	(0.0295)	(0.0450)
ln_EMA_pop	0.439***	0.337***
	(0.0446)	(0.0442)
Panel B: Nearest poleis		
	count of nearest poleis	count of nearest poleis
ln_IMA_pop	0.109***	0.0726**
	(0.0192)	(0.0289)
ln_EMA_pop	0.215***	0.149***
	(0.0282)	(0.0279)
Panel C: Polis indicator		
	polis indicator	polis indicator
ln_IMA_pop	0.130***	0.115***
	(0.0149)	(0.0198)
ln_EMA_pop	0.231***	0.185***
	(0.0191)	(0.0209)

Notes: This table varies the outcome variable in regressions otherwise identical to Table 4. Panel A copies, from Table 4, key coefficients for comparison. In Panel B we replace the outcome variable with a count of nearest poleis. In Panel C we replace the outcome variable with an indicator function of Poleis. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

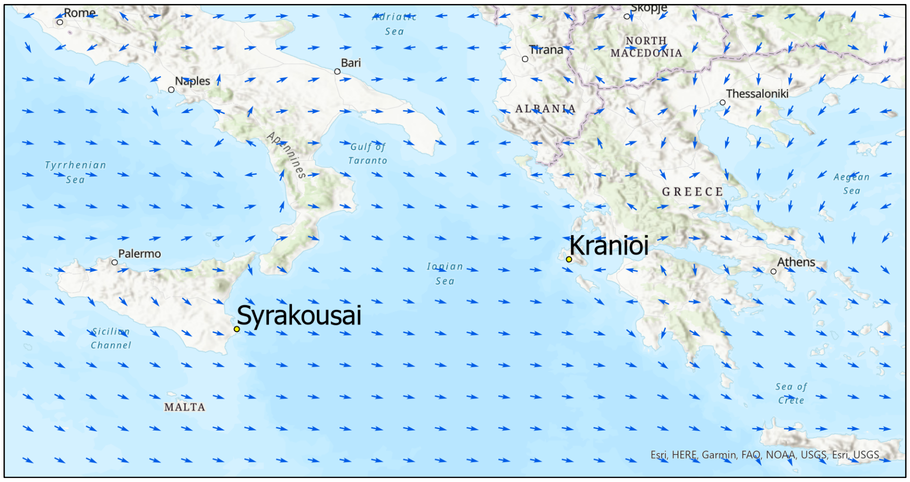


Figure A3: Example 1: Syrakousai and Kranioi

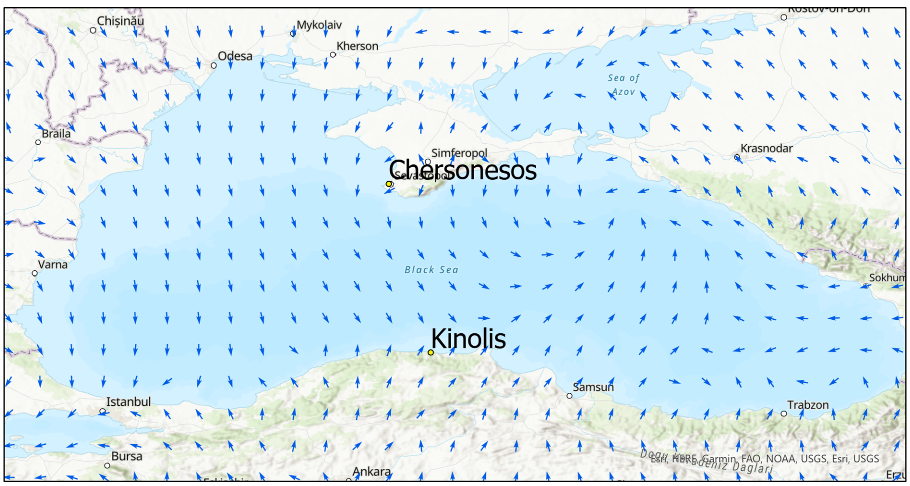


Figure A4: Example 2: Chersonesos and Kinolis

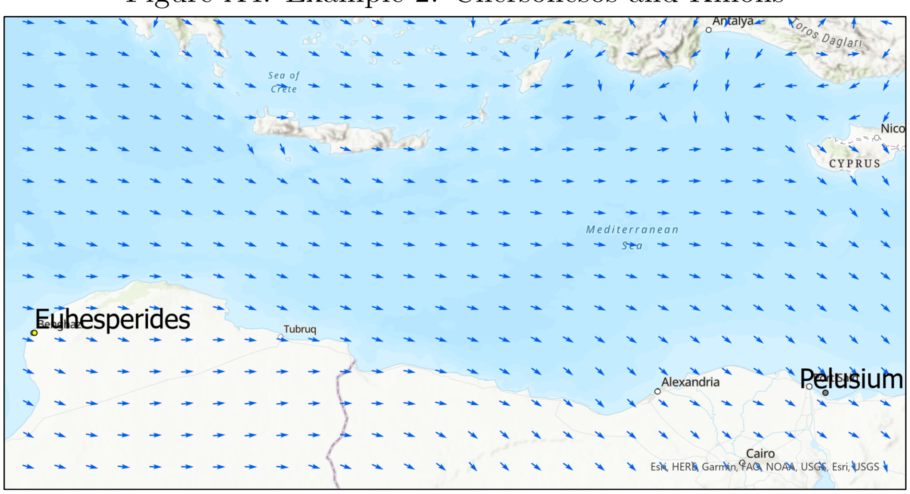


Figure A5: Example 3: Euhesperides and Pelusium

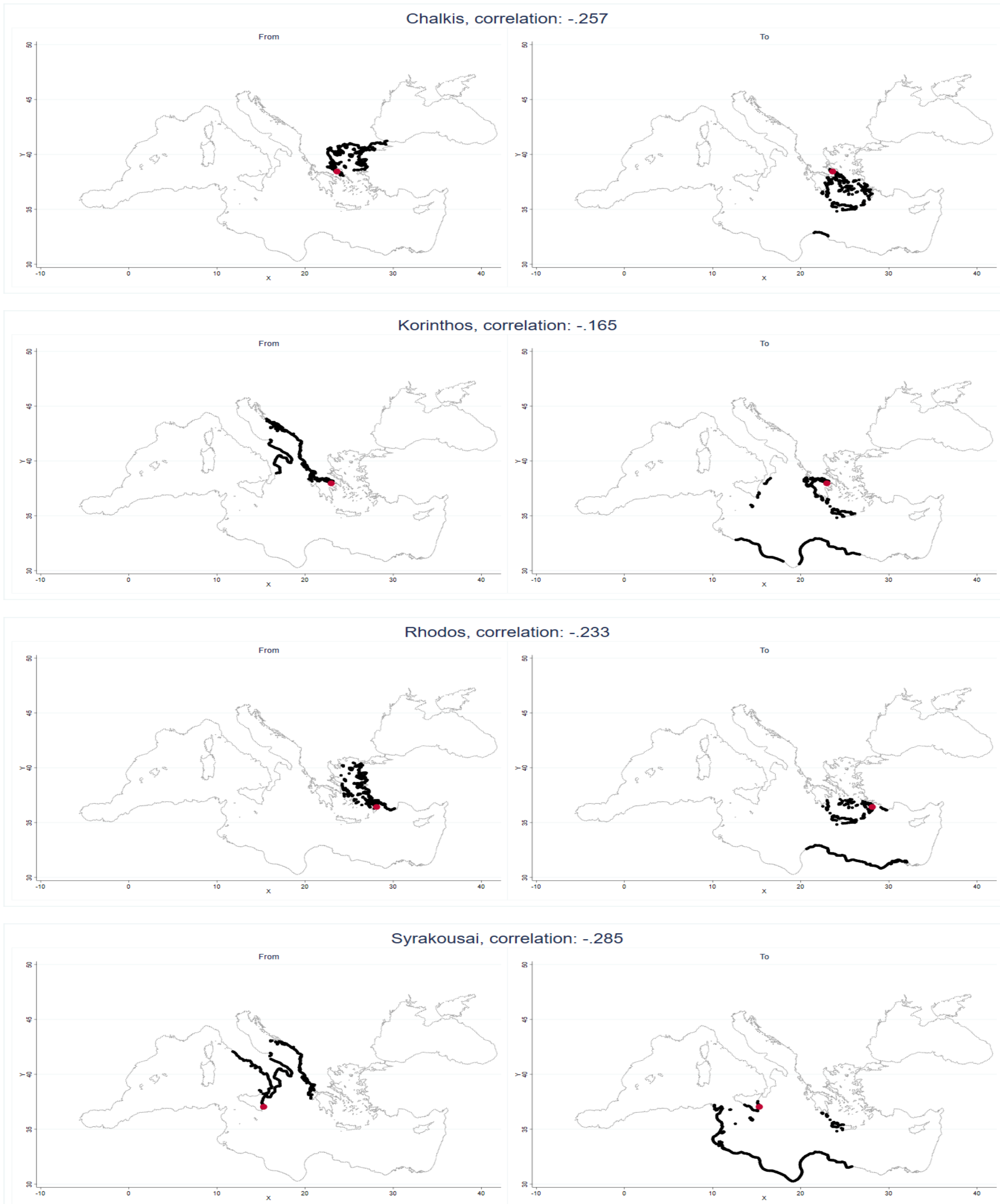


Figure A6: These maps show the 10 percent closest segment points *EMA* and *IMA* for four particular poleis. The plots also show the correlation between the *EMA* and *IMA* market access for segment points.