ASYMMETRIC TRADING COSTS AND ANCIENT GREEK CITIES[1]

Yuxian Chen*

Yannis M. Ioannides**

Ferdinand Rauch***

August 29, 2023

Abstract

Asymmetric transport costs arise when shipping times from point i to point j differ from shipping from point j to i. We show that such asymmetric transport costs predict distinct patterns of location in a class of models using Dixit-Stiglitz preferences, where advantage in import potential is more beneficial than that in export potential. We then study factors affecting the location of cities in Hellas, ancient Greece. Prevailing winds create an environment of asymmetric trade costs in ancient Greece. We show that predictions of these models are consistent with the location of ancient cities. We use 2SLS methods to assess the role of market access measures in determining urban settlements and find significant differences between the coefficients for export and import market access.

The authors have nothing to disclose.
Only resources of our universities, Tufts, Brown, and Heidelberg were used in this research.
JEL Codes: F10, N70, N90, R12.
Keywords: ancient urbanization, intercity trade, asymmetric trading costs, economic geography, spatial models, transportation networks, network centrality, systems of cities, market access, 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 Heidelberg. Germany. ferdinand.rauch@awi.uni-heidelberg.de

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. With strong wind patterns, ships may trade cheaply in one direction, yet return in a circular fashion. In this paper, we study this phenomenon of asymmetric trade costs. First, we show theoretically that asymmetric trade costs influence the attractiveness of locations in a standard Dixit-Stiglitz framework if symmetry is not assumed. We then take these predictions to a data set of 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. We present evidence in support of the predictions from the model. Ancient Greece is a perfect setting to study this question, given its reliance on sailing technology for trade in the period we study and its strong wind patterns. We demonstrate that asymmetric trading costs patterns have non-trivial implications on observed economic behaviour. We also demonstrate new facts regarding locational fundamental factors including seasonal variations in shipping costs (travel times) that determine the location of cities in ancient Greece.

What we know about classical Greece is associated primarily with its cities, indeed its city states, *poleis* [Hansen (2006)].[2] Classicists and historians have documented extraordinary patterns of trade, growth, and political innovations. Recent research by Ober (2015) and Bresson (2016) has 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 reported by Hansen and Nielsen (2004),[3] and have been recently been digitized. Those together with additional data from other sources are amenable, in principle, to quantitative analysis. According to Scheidel (2003), the broader Greek world experienced growth of population and improvement in living standards after the Dark Age.

We use theoretical frameworks from international trade and economic geography in order to investigate the origins of Hellenic cities and patterns of urbanization in the ancient Hellenic world, the ancient Hellas. We show that asymmetric impacts of trade costs are a general feature of simple models of trade. Combined with these data, we ask questions such as: Which geographic features can explain the location of ancient settlements?[4] Do natural factors matter more than distances to other cities? What is the role of seasonal variation?

We are addressing some of these questions by means of an exhaustive analysis for the poleis for which location and size information are available. We are investigating why the poleis, for which we have location information, emerged where they did in space relative to the universe of potential sites in the part of the world associated with the Hellenic world, that is, the entire coastline along the Mediterranean and Black Seas. Particular questions that can be examined with our expanded data set for all coastal segments of the Mediterranean and Black Seas would be whether cities that did emerge would have been predictable in terms of fundamentals.

1.1 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 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 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 studies that consider various aspects of Hellenic poleis data 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 and settling colonies, they transferred their superior technologies and institutions to local populations in new geographic areas with which they intermingled.[5] 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 [c.f. Henderson *et al.* (2018)]. They also employ the poleis data, a key ingredient of the research reported by the paper, for robustness. See *ibid.* Table 6. Clearly, the results imply that ancient settlements are associated with superior locations, however in their setting 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, trends in the regional presence of cereal, olive, and vine pollen that they estimate 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.[6] Scholars have established that classical Athens imported cereals from the Black Sea area [Bresson (2016)] in exchange for manufactures, while olive products were traded on a local or interregional scale.[7] 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 empirical identification of market forces in trade across regions of Greece is supportive when it comes to employing economic tools in studying ancient urbanization.

In a study related to ours, Bakker *et al.* (2021) 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 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 at the level of the entire world, by using a coarser definition of cells. Our study is closely related to this paper in terms of dataset structure, empirical approach and research question. The main insights from our paper are not part of their analysis, in particular the importance of asymmetric trade costs as well as our findings regarding the importance of seasonal differences.

The 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 impct on trade. However, we do note that the classics literature emphasizes the significance of the institutional structure of the Delian League for trade, and in particular, the harmonization of standards for weights and measures via the imposition of the Athenian *Coinage, Weights and Measures Decree* [see Kallet and Kroll (2020)] the group of allied poleis under the hegemony of Athens, and of other looser and lesser known associations of poleis in the form of regional federations.

Our paper also relates to an existing literature in trade that studies asymmetric trade costs. As Redding and Rossi-Hansberg (2017) conjecture, asymmetric shipping costs have 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 income differences between countries. There are also papers that endogenize trade costs and show that there are substantial differences of trade costs in trade data [Brancaccio *et al.* (2020); Ganapati *et al.* (2023)]. Closer related to our paper is an existing literature on gravity models and general equilibrium under asymmetric trade costs (Bergstrand *et al.* (2013); Melitz and Ottaviano (2008)]. What we contribute to this literature is a new, general theoretical framework showing how asymmetry matters for trade, and the empirical evidence that asymmetric trade costs contributed in a quantitatively important way to the location of Greek cities, and thus the geography of European history and economic development.

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 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 male members of the aristocracy, educated citizens of Hellenic poleis, did have such a perception of geography is evidenced by the historical record. From Herodotus we know that he himself 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?). Then a passage from Plato's *Phaedo*, 109a-109b, is quite suggestive:[8]

" '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,[9] whereby poleis 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)].

1.2 Summary of Findings

Since the main focus of this research is the importance of trade for polis sizes and patterns of urbanization in the ancient Hellenic world, our empirical analysis is motivated by a model of interpolis trade which rests on a standard model of international trade. Our empirical findings establish such importance by means of both centrality and market potential, that is, market access regressions which are motivated by our model. This model is appropriate because the ancient Greek cities, the Hellenic city-states, were in effect sovereign polities, which shared numerous characteristics including geographic ones. It is particularly noteworthy that the empirical analysis reveals that asymmetries in shipping costs are very important. Consistent with our model, the import potential is more important than export potential. However, such result is sensitive to the measure of trade potential. The result holds when trade potential is measured exactly in the form the model suggests (i.e., in terms of eigenvector centrality), but does not always hold when measured in alternative ways. We also run regressions with market access to urban settlements as the variables. Recognizing their endogeneity, we estimate it with 2SLS and find that with the market access to urban settlements, the export market access (which we refer to as *EMA*) is more important than import market access (*IMA*).

Specifically, we consider the coastline along the Mediterranean and Black seas as the universe of potential settlements. We divide the coastline into evenly spaced "segments", and use the midpoint of each segment for most of our spatial analysis. We conduct Poisson regressions, with the coastal segments as the unit of analysis, for settlement outcomes against trade potential to all other potential sites of settlements (that is, all segments) to test the prediction of our model about the role of asymmetric trade costs. We then also consider the impact of being close to other settled sites. Since the settlement outcomes are endogenous, the market access to settled sites are endogenous. To get consistent estimation, we use 2SLS with several different instruments, which we detail in the empirical section. The regressions typically yield statistically significant effects for market access to poleis in the vicinity of other segments, weighted by the respective shipping costs. We also explore the seasonality and the impact of seasonal variation, with the seasonal shipping cost data, and show that it is associated with larger settlements.

1.3 Outline of the Paper

The remainder of the paper presents next, section 2, a brief description of the economy of Ancient Greece (Hellas). Section 3 outlines a model of trading poleis (cities) under the Armington assumption using Dixit-Stiglitz preferences and uses its solution for populations and wages under the spatial equilibrium assumption of economic geography to motivate our empirical analysis. Section 4 discusses aspects of the data, with more detail on the poleis data set given in section 4.2, and on the segments data set given in section 4.3. Section 4.4 utilizes both data sets to define measures of urbanization, section 4.5 discusses the imputation of travel times. Sections 4.6 and 5.2 discuss alternative measures of trade potential. Our empirical results are presented in section 5. Section 5.1 tests the model predictions, namely the role of trade potential to the universe of potential urban settlements; section 5.3 discusses our findings regarding the importance of trade in explaining ancient urbanization using OLS methods; section 5.4 pursues, in particular, the importance of asymmetric trade costs and section 5.5 presents regression results with the market access to endogenously settled sites as explanatory variables. Last, section 5.6 highlights the role of seasonal variation in traveling costs. Section 6 concludes.

The basic model we propose below may be extended to additional directions in order to account for networks of alliances and how they might affect trade and therefore size of settlement patterns, emergence of new poleis, and political economy and diffusion of institutions. We plan to address these dimensions in future extensions of this work.[10]

2 The Economy of the Ancient Hellenic Poleis

The Ancient Greek cities, the Hellenic poleis, are instances of city-states (a contemporary term), of which there exist many different examples in world history [Woolf (2020), pp. 172-173], that have received a lot of attention for primarily cultural reasons. The term applies not only to the modern notion of a city but also the body of its citizens as well as its territory which functioned as city-states with their own self-government.[11] Such urban entities interacted with their immediate rural areas as well as other poleis. The *Inventory* [Hansen and Nielsen (2004)] covers a lot of historical, political and institutional detail for all 1035 poleis that may be identified to have existed during the archaic and classical Hellenic periods (650 BCE – 323 BCE) and are based on the archaeological and historical record. Some of them grew from original colonization by other poleis, and developed and progressively became Hellenized. Slavery was widespread but related legal protections differed widely across poleis. Not all of them are known to have issued coinage, but all of them traded extensively with each other, with non Greek-speaking city-states as well as distant areas.

The geographical reach of the market access of larger poleis is suggested by the historical evidence of "staggering series of luxury, strategic and basic commodities imported" by notable individuals [Kallet and Kroll (2020), p. 28-29]. Agricultural goods, especially cereals, olive oil and wine as well as such manufactured goods as textiles, tools, and artifacts were widely traded. Evidence of trade abounds in the archaeological and historical record. In addition, as we discussed earlier, Bonnier *et al.* (2020) use pollen data from a number of sites in southern Greece and Macedonia to study long-term vegetation change in these regions from 1000 BCE to 600 CE. Aside from a few large poleis, all other poleis were relatively small; population estimates are available only for only some of them and must be used with caution; see Ober (2015), p. 32, T. 2.1, Hansen and Nielsen (2004), Ober (2007), and Woolf (2020)].

In addition to Bresson (2016), Ober (2015), and Amemiya (2007) noted earlier, Woolf (2020) provides important context for all ancient cities within global urbanization. The latter dwells extensively on the interactions of Greek speaking and non-Greek speaking population centers. Unavailability of data, at least as systematically organized and documented as the poleis data from Hansen and Nielsen (2004) — not to mention that numerous sites have not been excavated, finds might not have been published, nor material collected in usable form — hampers broader studies of interactions between Hellenic and non-Hellenic city-states.

Central to this inquiry is understanding how the Hellenic system of cities has thrived on intercity trade. The evolution of the Hellenic urban world was a series of long episodes of growth of urban systems followed by decline and fall, especially by the time of the Macedonian conquest of Hellas and more [Scheidel (2003); Ober (2015)]. We take a broad view of the Hellenic world and examine, in particular, the extent in which urbanization in Hellas proper, essentially sites in the Aegean and what modern Greece could be explained in terms of trade with the entire system of poleis.

3 Theoretical Modeling

The rudimentary discussion of models of trade that follows fixes ideas, sets notation and motivates our empirical analysis. The literature distinguishes two principal cases [Allen, Arkolakis and Takahashi (2019), p. 399]: one, the "trade" model, with labor endowments in each location being exogenous and perfectly inelastic; and two, the "economic geography" 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)]. We appeal primarily to the second model to fix ideas and motivate our regressions. We also demonstrate that export and import market access matter in different ways in standard trade theory frameworks if trade cost symmetry is not assumed. Appendix A briefly mentions and makes a limited use of the first model.

Labor L_i is the only factor of production in 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 TFP A_i:

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

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 in the form of a CES function of $c_{ji}, j = 1, \ldots, J$, the quantities of the goods produced in site j and shipped 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:

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

Using (5) in (3) allows us to express indirect utility Ω_i in terms of relative prices only, $w_j/w_i, j \neq i$. This yields:

$$\Omega_i = u_i \left[\sum_k \left(\tau_{ki} \frac{w_k}{w_i} \frac{1}{A_k} \right)^{1-\sigma} \right]^{\frac{1}{\sigma-1}}. \tag{6}$$

If polis populations are given, the intercity trade equilibrium determines relative wages uniquely. If individuals were allowed 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 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. Following Allen and Arkolakis (2014), Theorem 1, p. 1094, we adapt their continuous space framework to our discrete space setting and show that the spatial equilibrium is fully characterized by the right and left eigenvectors associated with the Perron-Frobenius eigenvalue of the kernel of an underlying linear system, which is to be specified shortly.

Spatial equilibrium as an assumption is attractive because it yields testable predictions: wage rates, populations and utilities are all determined as functions of σ, site amenities and productivities, bilateral shipping costs and total population, that is in terms of exogenous variables. Imposing spatial equilibrium might sound as an extreme, but yet is quite plausible an assumption in the long run. The historical record suggests that poleis spawned other poleis by means of colonization and Hellenization of existing non-Hellenic settlements, as populations spread widely around the Mediterranean and the Black Sea regions. Indeed, the historical evidence attests to colonization and Hellenization being deliberate state actions in view of population growth and the pursuit of opportunities and geopolitical advantages.

We characterize spatial equilibrium by defining two auxiliary variables: $\tilde{x}$ and $\tilde{y}_i$:

$$\tilde{x}_i := w_i^{\sigma} L_i. \tag{7}$$

$$\tilde{y}_i := w_i^{1-\sigma}. \tag{8}$$

Employing $(\tilde{x}_i, \tilde{y}_i)$, and the additional definition

$$\tilde{A}_{ij} := u_j^{\sigma-1} \left(\frac{\tau_{ij}}{A_i} \right)^{1-\sigma}, \tag{9}$$

we obtain the systems of linear equations in $(\tilde{x}_i, \tilde{y}_i)$:

$$\tilde{x}_i = \sum_j \Omega_j^{1-\sigma} \tilde{A}_{ij} \tilde{x}_j; \tag{10}$$

$$\tilde{y}_i = \Omega_i^{1-\sigma} \sum_j \tilde{A}_{ji} \tilde{y}_j. \tag{11}$$

Under the Rosen-Roback or economic geography case, the terms are $\Omega_j^{1-\sigma}$ equalized across all poleis, and factor out of the summation in (10). Eq.'s (10–11) become:

$$\tilde{\mathbf{x}} = \bar{\Omega} \tilde{\mathbf{A}} \tilde{\mathbf{x}}; \tag{12}$$

$$\tilde{\mathbf{y}} = \bar{\Omega} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}. \tag{13}$$

Eq.'s (12) – (13) viewed as system of linear homogeneous equations in $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$, respectively, yield that $\bar{\Omega}$ is the Perron-Frobenius eigenvalue associated with the kernels of (12) and (13), respectively, which is common among them, because $\tilde{\mathbf{A}}^T$, the kernel of (13), is the transpose of $\tilde{\mathbf{A}}$, the kernel of (12). The only nonzero solutions for $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ are given by the Perron-Frobenius right and left eigenvectors of $\tilde{\mathbf{A}}$, a strictly positive nonsymmetric matrix. These solutions uniquely determine equilibrium real wages and populations up to scale, which together with total population determine populations at all sites. Note that the entries $\tilde{A}_{ij}$ combine the shipping costs from i to j with TFP at i and local amenities at j.

Finally, by manipulating the definitions of $(\tilde{x}_i, \tilde{y}_i)$, we may express the equilibrium wage rates and populations in terms of those auxiliary variables. That is, for population, in particular, we have:

$$L_i = \frac{\bar{L}}{\sum_j \tilde{x}_j (\tilde{y}_j)^{\frac{\sigma}{\sigma-1}}} \tilde{x}_i (\tilde{y}_i)^{\frac{\sigma}{\sigma-1}}. \tag{14}$$

The equilibrium is thus completely determined in terms of all exogenous variables, succinctly defined as:

$$\{\sigma, \bar{L}; \Upsilon\}, \quad \text{where} \quad \Upsilon := \{u_i, A_i, \tau_{ij} : i, j = 1, \ldots, J\}, \tag{15}$$

where the shipping costs $\mathcal{T} := [\tau_{ij}]$ may be conditional on seasons, $\mathcal{T}(season)$.

Spatial equilibrium is commonly invoked in studies with modern data. Equalization of utility may be a realistic assumption in our setting as well, particularly given that we study outcomes over a long period of time under the assumption of inter-city migration.[12]

We work with this model using imputed traveling times as proxies for the τ_{ij}'s. Amenities and productivities u_i, A_i are proxied by vectors of characteristics. From (9), the terms $\tilde{A}_{ij}$ depend on the productivity of the origin, A_i, and the consumer amenities at the destination, u_j.

Total population $\bar{L}$ is not observed in our data, nor are the populations $L_i's$ at each site. As we detail below, our data allow us to define a number of proxies for polis size, which we interpret as measures of the size of the economy of each polis. The data also include local productivity-related geographical characteristics at each geographical location, and travel times across sites, which proxy for trading costs. Our estimations are aimed at utilizing these data as we seek to explain urban development around the Mediterranean and Black Seas. We aim at explaining urbanization in the context of the geography of the Mediterranean and the Black Seas and in relation to local and global geographic characteristics that affect productivity and welfare. We do not, however, seek to explain the dependence of polis sizes on local spillovers, which improve local productivity, and on congestion, which reduce utility, nor their net effects. The Allen and Arkolakis (2014) framework is indeed amenable to this approach.

3.1 Testable Predictions

We bring to the data a number of testable predictions, which are based on our own adaptation of the trade literature, specifically on models of trade among independent poleis. These models are based on the Armington assumption with Dixit-Stiglitz Preferences, goods exchange subject to iceberg trade costs, and absence of monetary transactions. We present these predictions next.

1. With a given total population, $\bar{L}$, free trade subject to iceberg trade costs and free labor mobility, when spatial equilibrium prevails the utilities that individuals enjoy are equalized across all sites, $j = 1, \dots, J$, and site populations, L_i, are determined uniquely and satisfy $\bar{L} = \sum_i L_i$, and so are site wages, w_i. Site populations, given by Eq. (14), are expressed solely in terms of $(\tilde{x}_i, \tilde{y}_i)$, the right and left eigenvectors, defined by (12) and (13), of the matrix defined in (9) where: $\tau_{ij} \geq 1$, denote iceberg trade costs, that is τ_{ij} units of the good produced in site i must be shipped in order for 1 unit to arrive at site j; u_i denotes amenities entering utility of agents at site i; A_i denotes total factor productivity at site i (with labor being the only factor of production), and $\sigma > 1$ denotes the constant elasticity of substitution among all goods entering the individuals' utility function. Predicted populations according to (14) depend entirely on exogenous parameters $(\sigma; \bar{L}; \Upsilon)$. By taking logs, Eq. (14) yields a loglinear function of eigenvector centralities $(\tilde{x}_i, \tilde{y}_i)$:

$$ln(L_i) = ln\left(\frac{\bar{L}}{\sum_j \tilde{x}_j (\tilde{y}_j)^{\frac{\sigma}{\sigma-1}}}\right) + ln(\tilde{x}_i) + \frac{\sigma}{\sigma-1} ln(\tilde{y}_i), \tag{16}$$

   where the first term in the r.h.s is independent of i and thus absorbed by the intercept in regressions. We test this equation empirically in sections 5.3 and 5.4. Since the eigenvectors $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ are determined up to scale, the L_i's are scale-free. Note also that all amenity, productivity and shipping costs parameters enter $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ and thus affect each L_i.

   The empirical literature often posits that populations are functions of trade potential, which is proxied by measures of market access. That is, export (or alternatively producer, or outward) market access and import (or alternatively consumer, or inward) market access; they are denoted by EMA_i and IMA_i. They correspond conceptually to $\tilde{x}_i$ and $\tilde{y}_i$, respectively. We show below that they can also be justified from a network perspective as degree centralities; section 5.2.1. Although intuitively appealing modernizations of the Harris market potential [Harris (1954)], EMA_i and IMA_i are endogenous quantities. However, no model yields closed form solutions for the effects for such endogenous market potential. Therefore, our estimations of equations involving EMA_i and IMA_i rely on 2SLS methods; see section 5.2.1.

2. In the absence of trade, only the locally produced good is consumed, $C_{ii} = A_i$, and polis i real output and income (that proxy for polis size) is equal to $A_i L_i$, and utility per person is equal to $u_i A_i$. Therefore, all these quantities are functions of own characteristics u_i, A_i only and thus independent of all other elements of Υ, and notably of all shipping costs.

   This implies a straightforward identifiable restriction.[13] Reduced form regressions with proxies for urban settlements (or polis sizes) as dependent variables, and amenities and productivities at the corresponding segment as well as everywhere else and, in addition, shipping costs computed separately as via *EMA* and *IMA* as explanatory variables, that give significant results for global variables over in addition to local ones alone serve as evidence of trade. We test this prediction in section 5.3.

4 Data

4.1 Data Overview

We define a number of proxies for urban settlements at each segment s and regress them against observable characteristics of all segments, $s \in \mathcal{S}$ where $\mathcal{S}$ denotes the entire coast of the Mediterranean and Black Seas, as defined in terms of segments. We identify them as coastal segments and define them in more detail in the Data Appendix. There are 2,737 of them. We have distances and imputed travel times across all segments and a wealth of local crop suitability and climatic data, all of which included in our Segments Data. In addition, the segments are linked with a data set consisting of locations, their attributes and characteristics of Hellenic poleis, to be referred to as the Poleis Data Set. These data originate in the Stanford Polis Project, with additional data collected and coded by us.

We use a variety of regressions, instrumented as appropriate, and test their robustness by means of Poisson regressions, which account for the numbers of poleis as discrete variables and report results along these lines in section 5.5. We examine empirically the significance of seasonal fluctuations in shipping costs for settlement patterns by means of market access regressions with explanatory variables that are conditioned on seasonal shipping costs and, alternatively, on the annual standard deviation of shipping costs.

We discuss next in more detail the two principal data sets 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. We start with the archaic and classical poleis [Hansen and Nielsen (2004)].

4.2 The Poleis Data Set

The Poleis Data Set 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 about 1100 archaic and classical Hellenic cities by the *Copenhagen Polis Centre*.[14] From among the available variables of particular interest are the following: Size, the size of the polis territory in the archaeological record (see below for further details), region (namely historical region), 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.

Since the focus of analysis for this paper is the location choice of urban settlements, we rely heavily on such information. Thus, while the Poleis Data set has 1039 poleis in total, [15] 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 569 poleis for the final analyses.

The paper emphasizes economic interactions among archaic and classical Hellenic poleis, which are documented and curated by the *Inventory.* The poleis we are studying did interact extensively, like all cities of their era, with their hinterlands and other non-Hellenic poleis, as Woolf discusses extensively. In addition, all poleis in our data are rated in terms of their degree of Hellenicity. Among the original 1039 poleis, 820 are classified as purely Hellenic, 85 as mixed, and 44 mostly non-Hellenic; among the 571 poleis that we use, 478 are Hellenic, 48 mixed, 16 mostly non-Hellenic (and 29 of unknown degree of Hellenicity).

4.3 The Coastal Segments Data Set

An additional set of imputations and data collection involves 2,737 coastal segments that cover the entire coastline of the Mediterranean and Black Seas. We define these segment points by 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 data set 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.

Papers 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. We include a version of our preferred estimation that includes country fixed effects, which absorb differences in density of grid points at country level. These segments are fully identified in terms of geographical coordinates and linked to similar micro-geographic data as those of the poleis data set discussed above. Maritime travel times across all segments are also imputed and the details of these imputations are given in the Appendix, section C. Figure A2 in that section compares imputed with some historical travel times.

4.4 Measures of Urbanization

We measure urban settlement, that is, urbanization, by means of three variables that we define using both the Poleis and Coastal Segments Data sets. The first, an extensive margin measure, is the count of poleis settled at each coastal segment, $s \in \mathcal{S}$. 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 for all poleis that are within 10 km to the midpoint of a segment.

Polis size, however, is a complex matter, as Hansen and Nielsen discuss [*ibid.* 70–73]. Size, as available in the *Inventory*, is intended as a proxy for polis influence or territory, in other words, the size of the respective city-state. It is defined in *ibid.* as a categorical variable in reference to 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 4.2. 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. In recognition of the imprecision of the data, Hansen and Nielsen rely on inference that suffices to classify poleis in the above sparse categories. 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 assigns a few large poleis to size categories 6 and 7. Ober also suggests the following mapping from each polis size, as reported in the *Inventory*, to population: poleis of size 1, population 1,000; poleis of size 2, population 3,500; poleis of size 3, population 7,000; poleis of size 4, population 17,000; poleis of size 5, population 35,000; poleis of size 6, population 65,000; poleis of size 7, population 150,000 [*ibid.*, Table 2.1, p. 32].

Both intensive measures accords us a crucial advantage. Aggregating populations effectively over a segment territory obviates the need to define polis size and articulate heterogeneity in the density of settlement within the area defined as polis territory.

Given that trading costs 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 destructions over the ages, and accomodation 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 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 5.6 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.[16] In augmenting the data from the *Inventory* we add, in addition to the geological descriptions and agricultural suitability of the immediate environment of each polis, distances among sites and imputed travel times; see section 4.5.

4.5 Imputation of Travel Times

Wind is the central environmental factor in determining performance of ancient vessels. 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. Recent data could thus serve as proxies for ancient conditions. 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 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. Our imputations rely on GIS-based methods developed by Alberti (2018). Appendix, section C 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 is Gal *et al.* (2022), who emphasize the central importance of accounting for seasonality and currents, and also rely on GIS-based methods. We use both those other sources to conduct robustness exercises. However, 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. Imputations are discussed in the Appendix, section C. Figure A2 in that section compares imputed with some historical travel times.

4.6 Network-Based Centrality Measures of Trade Potential

Asymmetric trading costs is a notable feature of our data, whose significance we seek to explore empirically. In addition to the canonical market access approach we also work with network-based centrality measures taking advantage of their conceptual similarities. Eq. (14), or (16), expresses site populations in terms of $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, the right and left eigenvectors of the matrix defined in (9), which are functions of exogenous quantities.[17] Stachurski and Sargent (2022) refer to the right and left eigenvectors, defined by (12 – 13), respectively, as hub-based and authority-based eigenvector centralities. Given that the eigenvector centrality measures that the literature has considered are typically based on symmetric matrices, it is a distinct advantage of our approach that the matrix defined by (9) is asymmetric and therefore lends itself to a richer set of centrality measures. If we were to ignore the variations in A_i and u_j across sites and consider the case with $\sigma = 2$, (9) yields $\tilde{\mathbf{A}} = \left[\tau_{ij}^{-1}\right]$, which is defined entirely in terms of inverse trading costs across segments and is thus asymmetric.

One known issue with eigenvector centrality is that it can behave erratically with directed networks. It is for this reason that Kleinberg (1999) proposes two alternative centrality measures, namely authority and hub centrality, specifically for directed networks. The Kleinberg authority and hub centralities in non-matrix form are defined in terms of the elements of $\tilde{A} = \left[\tau_{ij}^{1-\sigma}\right]$ as follows:

$$\begin{aligned} \mathrm{Au}_i &= \sum_j \tau_{ji}^{1-\sigma} \mathrm{Hub}_j; \\ \mathrm{Hub}_i &= \sum_j \tau_{ij}^{1-\sigma} \mathrm{Au}_j. \end{aligned} \tag{17}$$

Authority centrality is conceptually akin to the gross value of imports (from "Hubs") by site i ; hub centrality is akin to the gross value of exports (to "Authorities") by site i.[18] By substituting the definition of *Hub* from the second equation in (17) into the first and rewriting in matrix form, and doing likewise from the first into the second we see that the Kleinberg centrality measures are defined as the Perron-Frobenius eigenvectors associated with the matrices $\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}$ and $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$, respectively.[19] Or, authority and hub centrality are the right and left eigenvectors of $\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}$. They, too, aggregate the impact of the entire urban system on urbanization at each segment. We compare their performance with those of eigenvector centralities of $\tilde{\mathbf{A}}$, and overall market access variables that we define in section 5.2.1 below.

5 Empirical Results

5.1 Basic Implications of the Model

Eq. (14) readily implies that segment populations may be expressed as log-linear functions of the eigenvector centralities $\tilde{x}_i$ and $\tilde{y}_i$, where σ, the elasticity of substitution among varieties, is unknown. We get around tedious structural estimation and obtain a rough sense of plausible values for σ by working as follows. For alternative values of $\sigma \in [1.1, 3.0]$, at increments of length 0.01, we first compute the associated $\tilde{x}_i$ and $\tilde{y}_i$, and then estimate Eq. (16) as a Poisson regression:

$$\ln L_i = a_0 + a_1 \ln(\tilde{x}_i) + a_2 \ln(\tilde{y}_i) + \mathbf{X'}_i \boldsymbol{\beta} + \varepsilon_i, \tag{18}$$

where $\mathbf{X}_i$ is a vector of site $i-$specific microgeographic characteristics. It is important to control for these as a (partial) remedy for the fact that our computations of $\tilde{x}_i$ and $\tilde{y}_i$ are based only on shipping costs across all sites and do not include the amenity and productivity parameters at all sites that enter matrix $\tilde{\mathbf{A}}$, defined in (9).

Our theoretical model predicts that coefficients (a_1, a_2) in Eq. (16) have the following values: $a_1 = 1, a_2 = \frac{\sigma}{\sigma-1}$. Figure 1 below plots, over the range of values of σ as defined above, the estimates $\hat{a}_1$ and $\hat{a}_2$ thus obtained and compares them with 1 and $\frac{\sigma}{\sigma-1}$, respectively. Interestingly, the plots of $\hat{a}_1$ and $\hat{a}_2$, as functions of σ, and the curves for their predicted values cross only once. We test whether $\hat{a}_1$ and $\hat{a}_2$ are different from their respective predicted values and find that with 95% probability we cannot reject that the estimated coefficients from regressions using the three different measures of L_i, defined in section 4.4, that $\hat{a}_1$ is different from 1 and $\hat{a}_2$ is different from $\frac{\sigma}{\sigma-1}$, for $\sigma \in [1.69, 1.74]$. These regressions are based on the Coastal Segments data.

5.2 Alternative Measures of Trade Potential

While the left and right eigenvector centralities readily follow from our theory, we also consider two additional, conceptually similar measures: one, the Kleinberg authority and hub centrality measures, which in network analysis are commonly believed to perform better than the eigenvector centrality for directed networks like ours; and, two, the canonical market access measure which is commonly used and easy to compute and interpret.

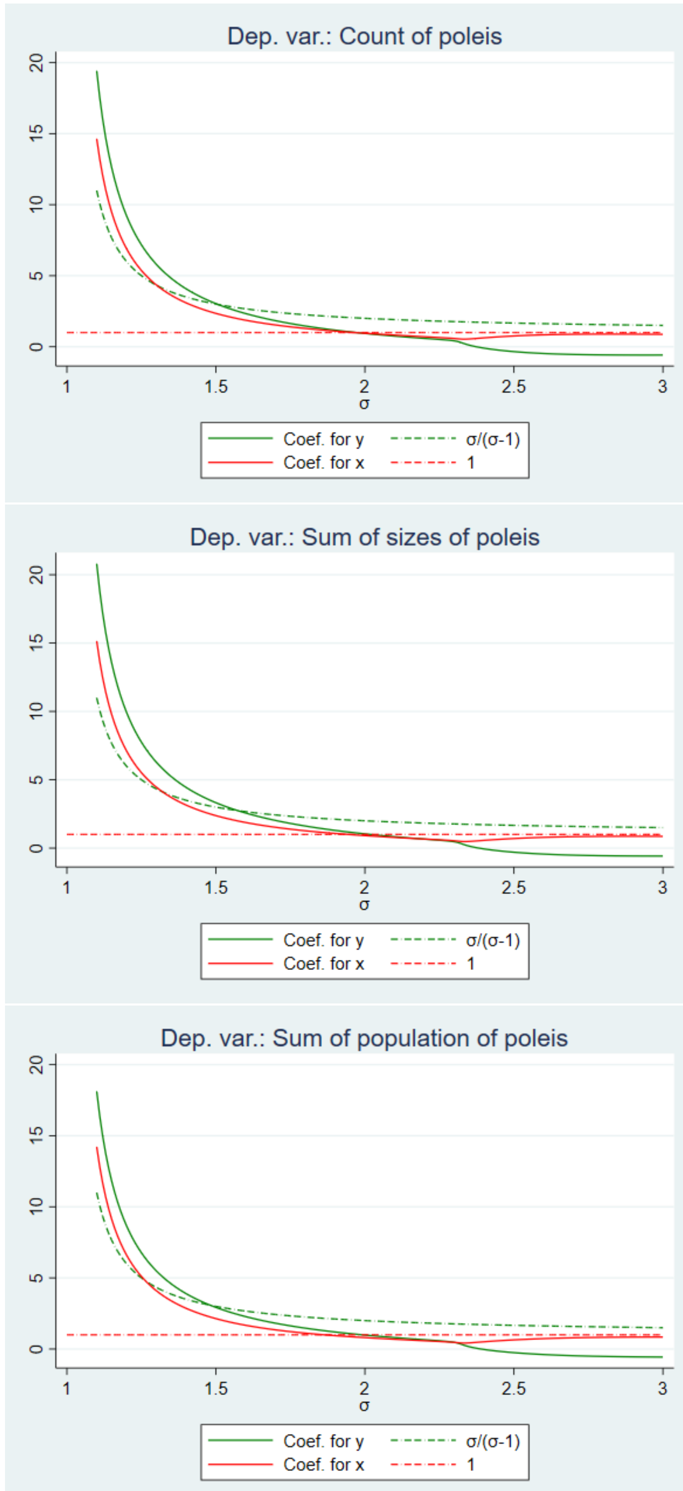


Figure 1: Estimates from Eq. (18), $\sigma \in [1.1, 3.0]$ (at 0.01 increments); $\tilde{A} := [inverse\ distances_{ij}]$

As both the Kleinberg centralities and the market access measures are directed, they are natural counterparts of the eigenvalue centralities and therefore allow us to test the robustness of our empirical findings regarding the main theoretical prediction. That is, that a segment's connections *from* other segments matters more than a segment's connection *to* other segments. In sum, we argue that Kleinberg authority centrality and import market access are conceptual counterparts of $\tilde{y}_i$, and Kleinberg hub centrality and export market access are conceptual counterparts of $\tilde{x}_i$. This claim is clarified further below in section 5.2.1.

5.2.1 Overall Market Access

The measure of trade potential discussed earlier is closely related to market access measures, which in our context would be defined at each segment as the sum of inverse distances to all other settlements. This is an endogenous magnitude: whether or not a segment is settled is endogenous. However, as we would like an exogenous measure, we define a segment's market access as the sum of inverse distances to all other segments, all of which are considered as potential sites for settlement.[20] That is:

$$\begin{aligned} \mathrm{EMA}_i &= \sum_j \frac{1}{\tau_{ij}}; \\ \mathrm{IMA}_i &= \sum_j \frac{1}{\tau_{ji}}. \end{aligned} \tag{19}$$

If $\sigma = 2$, the left and right eigenvector centralities coincide with import and export market access, respectively. If we restrict attention to the shipping costs part in the definition of $\tilde{A}$, Eq. (10) under spatial equilibrium becomes:

$$\tilde{x}_i = \bar{\Omega} \sum_j \frac{1}{\tau_{ij}} \tilde{x}_j \tag{20}$$

If we ignore the variation in $\tilde{x}_j$ in the r.h.s. of Eq. (20), it follows that $\tilde{x}_i$ is proportional to EMA_i, as defined by Eq. 19. In that case an alternative interpretation follows: $\tilde{x}_i$ denotes (weighted) *network degree* centrality.

As we saw in section 5.1, assuming $\sigma = 2$ may not yield the best fit. Thus, we define a

'Modified" MA measure by assuming an arbitrary value of σ:

$$\begin{aligned}\text{Modified\_EMA}_i &= \sum_j \tau_{ij}^{1-\sigma};\\ \text{Modified\_IMA}_i &= \sum_j \tau_{ji}^{1-\sigma}.\end{aligned} \tag{21}$$

We have experimented with both "Modified" MAs, using σ=1.71, and the conventional market access measures and report results with both measures.[21]

5.3 The Importance of Trade

We continue in like manner by estimating versions of Eq. (18) that allow us to examine the importance of trade in determining urbanization with alternative sets of proxies for trade potential. That is, we start with a basic specification with $\mathbf{X}_i$ as a vector of local productivity and amenity variables. We track the increase in R^2 as we use alternative proxies for trade. First, we estimate the model with the left and right eigenvector centralities, which readily follow from our theory. Second, we consider the two Kleinberg centrality measures, that is authority and hub centralities defined earlier. Third, we use measures of trade potential which is closely related to the popular market access measure; for details, see section 5.2.1 above.

Table 1, columns (1), (5) and (9) report the R^2 from regressions with the only local productivity variables as regressors. Columns (2), (6) and (10) report the R^2 from regressions with the local productivity variables and in addition the left and right eigenvector centralities; columns (3), (7) and (11) report the R^2 from regressions with the Kleinberg authority and hub centralities as proxies for trade; columns (4), (8) and (12) report the results with the import market access and export market access being added to the basic specification. Panel A reports the results for $\sigma = 1.71$ and Panel B for $\sigma = 2$.

The results with the left and right eigenvector centralities are as follows. In Panel A, we see modest increases in R^2 : from 0.163 to 0.241, with the log of count of poleis at a segment as the dependent variable (see columns (1) and (2)); from 0.142 to 0.228, with the log of the sum of polis sizes (see columns (5) and (6)). Compared with columns (1) and (2), the R^2 in columns (3) and (4) are smaller, suggesting our model is better at explaining the count of poleis (an extensive margin), probably because the measure of polis size is noisy. In columns (9) and (10), the dependent variable is the sum of polis populations at each segment, which is obtained from the size-population by size category from Ober (2015). The R^2 increases

	(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
Panel A: sigma=1.71												
r2	0.163	0.241	0.280	0.252	0.142	0.228	0.268	0.241	0.151	0.238	0.281	0.254
Panel B: sigma=2												
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

Table 1: Role of Trade: R2 from OLS Regressions

Notes: This table shows the increases in R^2 as different trade potential measures are added to the OLS regressions with only geographical variables as the explanatory variables, and have urban settlement measures (as indicated at the top of the table) as the outcomes. In columns (1), (5), and (9), only geographical variables (including precipitation, temperature, elevation (standard deviation), elevation (mean), malaria index, ruggedness, barley suitability, millet suitability, summer wheat suitability, and winter wheat suitability) are the explanatory variables; in columns (2), (6), and (10), logs of eigenvector centralities for both imports and exports are added along with the geographical variables as the explanatory variables; in columns (3), (7), and (11), logs of Kleinberg centralities for both imports and exports are added along with the geographical variables as the explanatory variables; in columns (4), (8), and (12), logs of (modified) market access (sum of inverse distances to the power of $\sigma - 1$) for both imports and exports are added along with the geographical variables as the explanatory variables. In panel A, all the trade potential measures are computed assuming $\sigma = 1.71$, while in panel B, all the trade potential measures are computed assuming $\sigma = 2$.

from 0.151 to 0.238 as we add the eigenvector centralities. This specification also works better than the log sum of size specification in terms of R^2.

The results with the Kleinberg authority and hub centralities (KC) are reported in columns (3), (7), and (11) of Table 1. 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 in part because they integrate export (outward) and import (inward) aspects of trade dependence.

We experimented with both Modified MA ($\sigma = 1.71$) and the conventional market access measure, MA. The former is reported in Panel A and the latter in Panel B, for both Table 1 and Table 3. In terms of explanatory power, the conventional MA has inferior performance to both those of eigenvector and Kleinberg centralities, as we can see from the R^2 from Panel B of Table 1. However, the performance of the Modified MA using $\sigma = 1.71$ (Panel A, Table 1) is better than that of the eigenvector centralities, but inferior to that of the Kleinberg centralities.

We also test for evidence of trade using the second prediction of section 3.1: in the

	(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: Dep var.: 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: Dep var.: ln(Sum of Polis Size +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: Dep var.: ln(Sum of Polis Population +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 2: Evidence of Trade: OLS coefficient for MA(x_j)

Notes: This table shows the OLS results for regressions that have urban settlement measures as the outcomes, and log of market access to geographical features (sum of inverse distances weighted by the specified geographical variable) 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. The geographical variable used to compute the market access for columns (1) – (10) are: precipitation, temperature, elevation (standard deviation), elevation (mean), malaria index, ruggedness, barley suitability, millet suitability, summer wheat suitability, and winter wheat suitability. Panel A, B, and C employ the same specifications but different measures for urban settlement as the dependent variable, as detailed in the table.

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

absence of trade, amenities and productivities at other segments and shipping costs should not matter once we control for the characteristics of the segment in question. We thus estimate the following equation:

$$\ln L_i = c_0 + c_1 \ln \left(\sum_{j \neq i} \frac{x_j}{d_{ij}} \right) + \mathbf{X_i'}\boldsymbol{\beta} + \varepsilon_i. \tag{22}$$

Here we aggregate characteristics of segments $x_j, j \neq i$, by a MA-style variable $\sum_{j \neq i} \frac{x_j}{d_{ij}}$. If there is no trade, c_1 should be zero. The regression results are reported in Table 2. Each column reports in turn different geographical variables as x_j, as indicated by the column heading. We can see that out of the 10 variables, 8 have significant c_1's. We conclude that trade is important for urbanization.

5.4 Role of Asymmetry in Shipping Costs under Alternative Measures for Trade Potential

Although we have already demonstrated the role of asymmetric shipping costs using the model, we would like to assess the role of asymmetric shipping costs with alternative empirical

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Count of Poleis			Sum of Size			Sum of Population		
	EV	KC	MA	EV	KC	MA	EV	KC	MA
Panel A: sigma=1.71									
ln(LEV/AU/IMA (y))	1.791***	1.198**	1.844***	1.979***	1.484***	2.159***	1.776***	1.424***	2.219***
	(0.424)	(0.488)	(0.539)	(0.411)	(0.468)	(0.523)	(0.381)	(0.487)	(0.503)
ln(REV/HUB/EMA (x))	1.503***	1.909***	1.915***	1.495***	1.743***	1.850***	1.322***	1.619***	1.720***
	(0.283)	(0.305)	(0.338)	(0.311)	(0.327)	(0.372)	(0.451)	(0.515)	(0.562)
P value(ln(y) $\neq$ ln(x))	0.489	0.270	0.903	0.232	0.686	0.587	0.365	0.821	0.493
Panel B: sigma=2									
ln(LEV/AU/IMA (y))	0.936***	0.187	0.891***	1.042***	0.525	1.158***	0.981***	0.657	1.397***
	(0.230)	(0.412)	(0.338)	(0.220)	(0.398)	(0.319)	(0.224)	(0.491)	(0.347)
ln(REV/HUB/EMA (x))	0.949***	1.479***	1.352***	0.916***	1.192***	1.252***	0.804***	0.972**	1.089***
	(0.169)	(0.298)	(0.229)	(0.182)	(0.317)	(0.254)	(0.266)	(0.471)	(0.375)
P value(ln(y) $\neq$ ln(x))	0.955	0.051	0.203	0.569	0.317	0.793	0.543	0.730	0.523
Obs	2737	2737	2737	2737	2737	2737	2737	2737	2737

Table 3: Role of Asymmetric Trade Costs: Poisson Regressions

Notes: This table reports Poisson regression results with urban settlement measures (as indicated at the top of the table) as the outcomes, and logs of trade potential measures (also indicated at the top of the table), 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 columns (1), (4), and (7), logs of the eigenvector centralities corresponding to imports and exports serve as the trade potential measures; in columns (2), (5), and (8), the logs of the Kleinberg centralities for both imports and exports are the trade potential measures; in columns (3), (6), and (9), the logs of modified market access (sum of inverse distances to the power of $\sigma - 1$) for both imports and exports are the trade potential measures. In panel A, all the trade potential measures are computed assuming $\sigma = 1.71$, while in panel B, all the trade potential measures are computed assuming $\sigma = 2$. The $p-$value comes from the Wald test for the equality of coefficients for logs of trade potentials for imports and exports.

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

measures of trade potential, namely the Kleinberg centralities and market access measures.

Although the authority centrality is conceptually closer to $\tilde{y}_i$ and the hub centrality to $\tilde{x}_i$, contrary to what we saw with the eigenvector centralities, here we obtain larger estimated coefficients for the hub centrality; see the results with the Poisson regression reported in columns (5) and (8) of Table 3.

Regarding the market access measure, our results vary depending on the choice of σ. With $\sigma = 1.71$, the Modified IMA always has a larger estimated coefficient than the Modified EMA; see columns (3), (6), and (9) of Panel A, Table 3. With $\sigma = 2$, that is with the conventional MA, we note that IMA does not always have a larger coefficient than EMA; see columns (3), (6), and (9) of Panel B, Table 3.

We conclude that the model's predictions regarding the role of asymmetric shipping costs are confirmed when the strict predictions of model are tested, that is by means of eigenvector centralities as in columns (1), (4) and (7), according to Eq. 16. When measured by means of Kleinberg centralities and Modified MA, the model prediction, namely importing matters more than exporting, is not robust.

5.5 The Effects of Export and Import Market Access

It may be argued that Eq. (22) does not fully express market access for each polis. As Appendix, section A, elaborates, there are two principal forces affecting welfare at each site. Individuals at a particular segment benefit from the demand for their output, that is for their labor, from the entire urban system via the export market access. They also benefit from a greater variety of goods they consume, the import market access. In the model we have worked with so far the number of goods is given by the Armington assumption. However, in more general models, like those reviewed by Redding and Rossi-Hansberg (2017) and Redding (2022), the number of goods vary by the size of each site. Then, the greater the size of a site from which a given site is importing, the greater the variety of goods it is producing and the lower the associated price index. We thus posit the following equation which generalizes Eq. (22):

$$\ln L_i = b_0 + b_1 \ln \left(\sum_j \frac{L_j}{\tau_{ij}} \right) + b_2 \ln \left(\sum_j \frac{L_j}{\tau_{ji}} \right) + \mathbf{X_i'}\boldsymbol{\beta} + \varepsilon_i. \tag{23}$$

The parameters we are interested in are b_1 and b_2, while $\mathbf{X}_i$ is controlled for. The motivation for the above equation in Appendix, section A, suggests that both b_1 and b_2 should be positive. Again, we use three measures for L_i, for both the dependent variable and the terms $\sum_j \frac{L_j}{\tau_{ij}}$and $\sum_j \frac{L_j}{\tau_{ji}}$ on the r.h.s. of Eq. (23). $\mathbf{X_i}$ is the same set of local geographical variables as before.

The challenge of estimating Eq. 23 is that the terms $L_j's$ in the terms $\sum_j \frac{L_j}{\tau_{ij}}$ and $\sum_j \frac{L_j}{\tau_{ji}}$ in the r.h.s of (23) and the corresponding sets over which the sums are defined are endogenous. To address the endogeneity, we construct instruments as follows. First, we ignore the variation in $L_j's$, and use $\sum_j \frac{1}{\tau_{ij}}$ as the instrument for $\sum_j \frac{L_j}{\tau_{ij}}$. Second, we obtain predictions for the $L_j's$, $\hat{L_j}'s$, and use them to construct the instrument. That is, we run a Poisson regression of L_i against the vector of local productivity variables $\mathbf{X}_i$ and obtain the predicted $\hat{L_j}$. We also obtain $\hat{L_j}$ from a Poisson regression of each L_i against the local amenity and productivity variables and the centralities, as suggested by (16). We use $\sum_j \frac{\hat{L_j}}{\tau_{ij}}$ as the instrument for $\sum_j \frac{L_j}{\tau_{ij}}$. We report in Appendix B results with a third method which originates in the network econometrics literature. Following Bramoullé et al. (2009), $\boldsymbol{G^2x}$

can be used as the instrument for $\boldsymbol{Gy}$, where $\boldsymbol{G}$ denotes the adjacency (interaction) matrix of the network.[22] In our case, $\boldsymbol{G}$ is defined as $[\frac{1}{\tau_{ij}}]$. and with y_i being L_i, $\boldsymbol{Gy}$ is just $\sum_j \frac{L_j}{\tau_{ij}}$. Thus, we can take the local amenity and productivity variables $\mathbf{X}$ to compute $\boldsymbol{G^2X}$ and use it as the instrument for $\boldsymbol{Gy}$ or $\sum_j \frac{L_j}{\tau_{ij}}$. Since the original method is intended for level-level regressions, for this IV method we estimate a level-level regression and include $\boldsymbol{GX}$ as the control variable following Bramoullé et al. (2009). We report the results in Appendix B because of the following reasons. One, there is a difference in specifications; and two, the network-based IVs are weaker.

The OLS and 2SLS regressions results are reported in Table 4. Column (1) reports the OLS results which show that with any of the measures for L_i, exports always have a larger effect than imports. To correct for the potential bias with OLS regressions, Columns (2) – (7) report the 2SLS regressions results. In column (2), we use the first IV method, namely ignoring $L_j's$ and using $\sum_j \frac{1}{\tau_{ij}}$ and $\sum_j \frac{1}{\tau_{ji}}$ as the instruments. The instruments in columns (3) – (7) use the second method, namely $\sum_j \frac{\hat{L_j}}{\tau_{ij}}$ and $\sum_j \frac{\hat{L_j}}{\tau_{ji}}$ serve as the instruments, where $\hat{L_j}$ stands for the predicted value from a Poisson regression of the dependent variable on geographical variables only, geographical variables plus eigenvector centralities (with $\sigma = 1.71$), geographical variables plus eigenvector centralities (with sigma=2), geographical variables plus Kleinberg centralities (with $\sigma = 1.71$), and geographical variables plus Kleinberg centralities (with $\sigma = 2$), reported in columns (3) – (7) respectively. When we compare with OLS, we generally see smaller coefficients with 2SLS regressions, probably reflecting endogenous clustering of urban settlements. However, the coefficients are still positive, and in terms of the relative magnitudes, export market access is more important than import market access. The differences are statistically significant and are robust across the different IVs. Appendix B reports the Poisson and IV Poisson regression results for the same specification. We note that the results are qualitatively similar, although the difference between the coefficients for lnIMA_L and lnEMA_L are less significant. We also report in Appendix B the level-level specification results with network IVs. Some of the instruments are weak, but whenever the instruments are strong, we always see a larger coefficient for EMA_L than IMA_L.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	2SLS					
		NoL	L:Geo	L:Geo+EV(s=1.71)	L:Geo+EV(s=2)	L:Geo+KC(s=1.71)	L:Geo+KC(s=2)
Panel A: L: 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: Sum of Polis Size							
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: Sum of Polis Population							
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
Obs	2737	2737	2737	2737	2737	2737	2737

Table 4: 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 distances to/from all segments (effectively ignoring the numerators $L_j's$) 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 $L_j's$ as the numerators) as the instruments. In the regressions reported in columns (3) – (7), the $L_j's$ 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.6 The Effects of Seasonal Variation of Shipping Costs

Seasonal variation of shipping costs implies seasonal variation in prices faced by individuals and firms. In partial equilibrium settings, indirect utility and profit functions are quasi-convex in prices, suggesting that variations in prices is welfare improving. However, that is not the case in general equilibrium, where seasonal variation implies variation in the value of endowments.[23] Although theory does not offer specific predictions regarding the impact of seasonal variations in shipping costs, it is interesting to see their impact on urbanization.

We explore the effect of seasonal variation by using measures of dispersion of the seasonal travel cost data as additional explanatory variables in regressions along the lines of those reported in Table 4. We compute first the coefficient of variation for the market access variables:

$$CV(MA)_i = \frac{\sqrt{1/3 * \sum_s (MA(t)_i - \overline{MA_i})^2}}{\overline{MA_i}}, \tag{24}$$

where $MA(t)_i$ is the overall market access (sum of inverse distances) 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 Eq. (24) accordingly. For our second definition, instead of $\overline{MA_i}$ we use market access computed at the yearly average travel costs:

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

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:

$$CV(\tau)_i = \sum_{k \neq i} \frac{\sqrt{1/3 * \sum_s (1/\tau(t)_{ki} - \overline{1/\tau_{ki}})^2}}{\overline{1/\tau_{ki}}}, \tag{25}$$

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 $1/\tau_{ki} = \tau_{ki}^{1-\sigma}$ for $\sigma = 2$. $\overline{1/\tau_{ki}}$ denotes the mean over the four seasons. Last, we also compute a modified version of (25) where instead of $\overline{1/\tau_{ki}}$ we use the travel cost based on yearly average wind data. That is,

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

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 sites. We refer to them as settlement CVs; for example, for the first definition of the CV in (24), the respective settlement CV involves $MA(s)_i$ as the sum of inverse distances to settled segments only, rather than to all segments. Similarly, for definition (25), the respective settlement CV would sum only over settled segments.

Table 5 reports results of regressions along the following equation:

$$\ln L_i = c_0 + c_1 \ln ICV_i + c_2 \ln ECV_i + c_3 \ln IMA(year)_i + c_4 \ln EMA(year)_i + \mathbf{X'}_i\boldsymbol{\beta} + e_i, \quad (26)$$

where the dependent variable L_i is either count of poleis, sum of poleis size, or sum of populations, and $\mathbf{X}_i$ is the same set of geographical variables as those used in previous regressions. We control for $\ln IMA$ and $\ln EMA$ to separate out the effect of seasonal variation from those of average market access. We estimate the equation with OLS when the right hand side CVs are the exogenous CVS, and with 2SLS when the right hand side CVs are the settlement CVs, using the exogenous CVs as instruments.

The regression results reported in Table 5 show very similar results with the exogenous CVs (columns (1) to (4)) and with the settlement CVs. The seasonal variation in travel costs 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 that for importing, which is consistent with the relative magnitudes of imports and exports that we discuss earlier in section 5.5. Last, the results for the settled segments are quantitatively similar to the respective ones for the universe of segments. As we discuss earlier, the failure of theory to offer specific predictions on the effects of seasonal variations make our findings quite provocative. However, as we conjectured in section 4.4, larger settlement sizes associated with more pronounced seasonal fluctuations may mitigate risk. Specifically, eschewing a formal derivation, we conjecture 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)
	Overall CV (OLS)				Settled CV (2SLS)			
	CV(MA)	CV(MA')	CV(τ)	CV(τ')	CV(MA)	CV(MA')	CV(τ)	CV(τ')
Panel A: L: 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: Sum of Polis Size+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: Sum of Polis Population+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
lnMA	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 5: Effect 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 coefficients of variation for travel costs over four seasons for both imports and exports, as well as log of sum of inverse distances 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 coefficient of variation (CV), as defined in section 5.6. The CVs in columns (1) – (4) are computed using travel cost 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusions

This paper examines the importance of trade for the sizes of Hellenic poleis, notable early urban settlements, and patterns of urbanization more generally in the ancient Hellenic world. The research pays particular attention to the consequences of asymmetries in transport costs, which are due to prevailing winds. 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 a vector of productivity and amenity characteristics.

Our empirical analysis is motivated by a standard model of international trade which demonstrates that asymmetries in trade costs affect the sizes of poleis (cities). The strict version of the model suggests that settlement outcomes are log linear functions of the eigenvector centralities associated with the spatial equilibrium model. We show that the model holds over a reasonable range of values of the elasticity of substitution among the varieties produced at the various sites and consumed everywhere. When the eigenvector centralities are used as regressors, the estimates are consistent with the predictions of the model, namely that import potential is more beneficial than export potential. However, this prediction is not confirmed when alternative measures of trade potential are used, such as the Kleinberg centralities and import and export market access. The results are less robust in those cases. We also explore the impacts of market access to actual settlements. By using a variety of instruments, we obtain a robust result that the export market access to settlements is more important than import market access to settlements. In addition, we document that seasonal variations in travel costs are beneficial for the emergence of early poleis and imply that sizes of settlements function as insurance against such variations.

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, that it, the ancient Greek cities 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.

There exist numerous other important factors that should have affected Hellenic urbanization that 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.

Notes

[1]We acknowledge helpful comments from a number of scholars without implicating them, especially from Costas Arkolakis, John Kroll and Josiah Ober, but also from Yann Bramoullé, Alain Bresson, Aureo De Paula, Yannis Evrigenis, Bruce Hitchner, Misha Kilmer, Hon Ho Kwok, Jacques Melitz, James Murphy, Theodore Papageorgiou, Herakles Polemarchakis, Harilaos Psaraftis, Diego Puga, Peter Van Alfen and seminar participants at various institutions. We also acknowledge help from Stelios Michalopoulos, who was helpful with the construction of the data on sailing costs among poleis, and from Nick Reynolds, who actually constructed the sailing costs data set. We are grateful to Dr. Takvor Sukissian of the Hellenic Center for Marine Research, and Christopher Barnett, Carolyn Talmadge, Uku-Kaspar Uustalu and the other members of the Tufts Data Lab team for the construction of the data set for sailing times across the coastal segments of the Mediterranean and the Black Seas. We benefited from excellent research assistance by Joseph Chance, Xu Dong, Gengbin Piao and James Plosser.

[2]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 refers to poleis, rather than cities.

[3]These data have been digitized originally by Josiah Ober and his team of Stanford scholars, as well as by ourselves working from Hansen and Nielsen (2004).

[4]Hicks (1969) argued that the Mediterranean geography was conducive to urban development in the broadest of senses, economic, intellectual and political, long before such issues were addressed quantitatively. More recently, Matsuyama (2017) conjectures in like manner.

[5] For example, this explains the Phoenician origin of the Greek alphabet; relatedly, see Chen and Ioannides (2022). 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 Etruscans, through whom it found its way to the Romans and came to be known as the Latin alphabet.

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

[7]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.

[8]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."

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

[10]Issue of coinage by poleis is also important but taken up in a separate investigation; see Chen and Ioannides (2022).

[11]According to 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*)." Aristotle first leaves out the third of three elements that comprise the modern juristic idea of a state, territory. But later on he considers the place where citizens live as a most important characteristics of a city.

[12]Abulafia (2011) comments on the multi-cultural nature of port cities from the time of the Phoenicians.

[13] Although climatic conditions across broad regions in the area of study are likely correlated, they do exhibit considerable local variations, especially with respect to microclimates affecting navigation [Beresford (2012)]. Therefore, local climatic conditions do not necessarily proxy for those at other sites.

[14] Some of the data were digitized by Josiah Ober and his team of Stanford and Oxford scholars

[http://polis.stanford.edu/].

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.

[15]We also did other data processing, including dropping duplicated observations and non poleis sites (such as regions and modern cities).

[16]We are naturally aware of potential climatic changes over more than 2500 years, but there exist very little detailed paleoclimatic data for this region of the world. We argue that our inference holds if the entire geographic area where our cities are located suffered similar climatic shocks. After inspection of the data, we sought to correct some inconsistencies, especially regarding some GPS coordinates, and to augment the data on account of missing information on some poleis. That investigation revealed that for some of the poleis missing information might not be recoverable, as some data 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.

[17] The importance of eigenvector centrality measures is increasingly acknowledged by the economics literature; see Bloch *et al.* (2021), and most recently Stachurski and Sargent (2022), Ch. 1, section 1.4.3.4.

[18]Stachurski and Sargent (2022) interpret left and right eigenvector centrality of production networks as authority and hub centralities which correspond to important suppliers, upstreamness, and to important demanders, downstreamness.

[19]Chen and Ioannides (2022), section 3.5, make similar use of eigenvector centrality measures in investigating their impact on the diffusion of coinage and script adoption across Hellenic poleis.

[20]This approach is similar to the definition of *connectivity* by Bakker *et al.* (2021). That is, a segment's connectivity is the number of other segments that can be reached within a specified shipping distance.

[21] While our MA measures emerge naturally in our case, they may be seen as a generalization of node centrality with σ playing the role of "tuning" parameter which produces smaller contributions from trade with more distant sites. See Opsahl *et al.* (2010).

[22] In network theory, the square of the adjacency matrix defines the set of second neighbors. For weighted networks like ours, 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.

[23] We thank Herakles Polemarchakis for help in clarifying this argument. In general equilibrium, the dependence of incomes (which are in effect 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.

7 References

Abulafia, David. *The Great Sea: A Human History of the Mediterranean.* Oxoford: 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, Costas Arkolakis, and Yuta Takahashi. 2019. "Universal Gravity." *Journal of Political Economy.* 128(2):393-433.

Allen, Treb, and Dave Donaldson. 2020. "Persistence and Path Dependence in the Spatial Economy." https://www.nber.org/system/files/working_papers/w28059/w28059.pdf.

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

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. 2022. “The Diffusion of Epichoric Scripts and Coinage in the Ancient Hellenic World.” Working Paper. Tufts University. Rev. Aug. 19, 2023.

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.

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

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

Ganapati, Sharat, Woan Foong Wong and Oren Ziv, “Entrepôt: Hubs, Scale, and Trade Costs”, University of Oregon, Working Paper, 2023.

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.

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

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

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.

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

Melitz, Marc J., and Gianmarco I. P. Ottaviano. 2008. “Market size, Trade, and Productivity.” *The Review of Economic Studies.* 75.1:295–316.

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.

Opsahl, Tore, Filip Agneesens, and John Skvoretz. 2010. "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths." *Social Networks.* 32:245–251.

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

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

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.

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 Motivation for Eq. (23)

If site populations were known, we may characterize the equilibrium of the model by means a system of equilibrium equations expressed in terms of the wage rates as unknowns, $\mathbf{w} = (w_1, \ldots, w_i, \ldots, w_J)$. Site i's sales to j, typically referred to as trade flows from i to j, are given by:

$$X_{ij} = \left(\frac{\tau_{ij}}{A_i u_j} \right)^{1-\sigma} w_i^{1-\sigma} \Omega_j^{1-\sigma} w_j^{\sigma} L_j. \tag{27}$$

For trade equilibrium, first, labor incomes, that is payments to labor in every city equal total sales, both from exports and local ones,

$$w_i L_i = \sum_j X_{ij}; \tag{28}$$

second, trade is balanced, that is, site i's share of total income going to consumption, $w_i L_i$, is spent on imports, which in turn implies,

$$w_i L_i = \sum_j X_{ji}. \tag{29}$$

Working with (27) and (29) and simplifying yields

$$w_i L_i = \sum_j \frac{w_i^{1-\sigma} \left(\frac{\tau_{ij}}{A_i} \right)^{1-\sigma}}{\sum_k \left(\frac{\tau_{kj}}{A_k} \right)^{1-\sigma} w_k^{1-\sigma}} w_j L_j. \tag{30}$$

Intuitively, the numerator of each term in the summation of the r.h.s of (30) denotes nominal demand by site j for site i output; the denominator denotes the price index of site j. Therefore, the first term in logs on the r.h.s. of Eq. (23), the *From* term, reflects the nominal demands, which do depend on the sizes of all other sites. Dividing by the distance reflects the term $\tau^{1-\sigma}$ for $\sigma = 2$. The contribution of this term should be positive. The second term, the *To* term, does not explicitly depend on the sizes of the respective sites, because all of the goods produced, which are as many as the sites (by the Armington assumption), are consumed by all sites. However, that is an artifact of the assumption that the range of the varieties is not endogenous and fixed by the number of sites. Dividing by the distance follows for the same reason as above. In general, with endogenous number of varieties in each site, the price index would be decreasing in the variety of goods imported and thus in their range, which in turn depends on the size of the site, and therefore the effect of the *To* term should also be positive.

Rationalizing fully this prediction requires more general models, like those reviewed by Redding and Rossi-Hansberg (2017) and Redding (2022). We eschew such modeling and rely instead on the following intuition. The number of goods vary by the size of each site. Then, the greater the size of a site, the greater the variety of goods it is producing and exporting and the lower the associated price index and the greater the welfare of the sites that are importing them.

Following Allen and Arkolakis (2015), Lecture 5, we may prove existence of equilibrium as follows. We rewrite (30) to define excess demand functions:

$$Z_i(\mathbf{w}) := \frac{1}{w_i}\left[\sum_j \frac{w_i^{1-\sigma}\left(\frac{\tau_{ij}}{A_i}\right)^{1-\sigma}}{\sum_k \left(\frac{\tau_{kj}}{A_k}\right)^{1-\sigma} w_k^{1-\sigma}} w_j L_j - L_i\right], j = 1, \ldots, J. \tag{31}$$

Following Mas-Colell, Whinston and Green (1995), Proposition 17.C.1, Allen and Arkolakis (2015) establish that there exists a unique strictly positive vector $\mathbf{w}$ that solves the system of Eq.'s (31) for $Z_i(\mathbf{w}) = 0, j = 1, \ldots, J$. Once wages are solved for, nominal and real incomes follow. However, this approach makes it hard to have a sense of the determinants of the wage rates, other that they depend on all parameters and the given populations $(L_1, \ldots, L_J)$. This is clearer in the case of spatial equilibrium. Given that, as we argue in the main paper, it is not a farfetched assumption we rely on it extensively.

B Additional Tables

B.1 Poisson and IV Poisson Regression results for Eq. (23

)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Poisson	IVPoisson					
		NoL	L:Geo	L:Geo+EV(s=1.71)	L:Geo+EV(s=2)	L:Geo+KC(s=1.71)	L:Geo+KC(s=2)
Panel A: L: Count of Poleis							
ln(IMA_L (y))	0.735***	0.355**	0.409***	0.418***	0.409***	0.435***	0.426***
	(0.125)	(0.169)	(0.153)	(0.155)	(0.155)	(0.155)	(0.156)
ln(EMA_L (x))	0.873***	0.638***	0.778***	0.774***	0.773***	0.775***	0.781***
	(0.113)	(0.111)	(0.111)	(0.110)	(0.110)	(0.111)	(0.111)
P value(ln(y) $\neq$ ln(x))	0.523	0.233	0.121	0.132	0.123	0.151	0.134
Panel B: L: Sum of Polis Size							
ln(IMA_L (y))	0.770***	0.514***	0.495***	0.488***	0.481***	0.502***	0.493***
	(0.124)	(0.159)	(0.138)	(0.143)	(0.143)	(0.143)	(0.144)
ln(EMA_L (x))	0.778***	0.540***	0.663***	0.690***	0.685***	0.694***	0.695***
	(0.134)	(0.130)	(0.129)	(0.128)	(0.127)	(0.128)	(0.127)
P value(ln(y) $\neq$ ln(x))	0.970	0.916	0.477	0.398	0.392	0.423	0.397
Panel C: L: Sum of Polis Population							
ln(IMA_L (y))	0.920***	0.674***	0.559***	0.503***	0.502***	0.519***	0.516***
	(0.148)	(0.184)	(0.144)	(0.149)	(0.151)	(0.153)	(0.158)
ln(EMA_L (x))	0.907***	0.475**	0.603***	0.668***	0.662***	0.689***	0.690***
	(0.181)	(0.216)	(0.210)	(0.204)	(0.202)	(0.202)	(0.200)
P value(ln(y) $\neq$ ln(x))	0.960	0.554	0.885	0.586	0.596	0.582	0.575
Obs	2737	2737	2737	2737	2737	2737	2737

Table A1: Role of Asymmetric Trade Costs and Endogenous Market Access: Poisson and IV Poisson Regressions

Notes: This table reports Poisson and IV Poisson 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 Poisson result. Column (2) is the IV Poisson result with sum of inverse distances to/from all segments (effectively ignoring the numerators $L_j's$) in forming the instrument. Columns (3) – (7) are the IV Poisson results with market access weighted by predicted urbanization measures (effectively using predicted rather than actual $L_j's$ as the numerators) as the instruments. In the regressions reported in columns (3) – (7), the $L_j's$ 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.

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

B.2 Network IV results for the level-level specification of (23)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	prec	temp	elev(std)	elev(mean)	malaria	rugg	Crop:B	Crop:M	Crop:SW	Crop:WW
Panel A: L: 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: Sum of Polis Size										
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: 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 A2: Role of Asymmetric Trade Costs and Endogenous Market Access: 2SLS with network IVs

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 Sailing Times across Coastal Segments of the Mediterranean and Black Seas

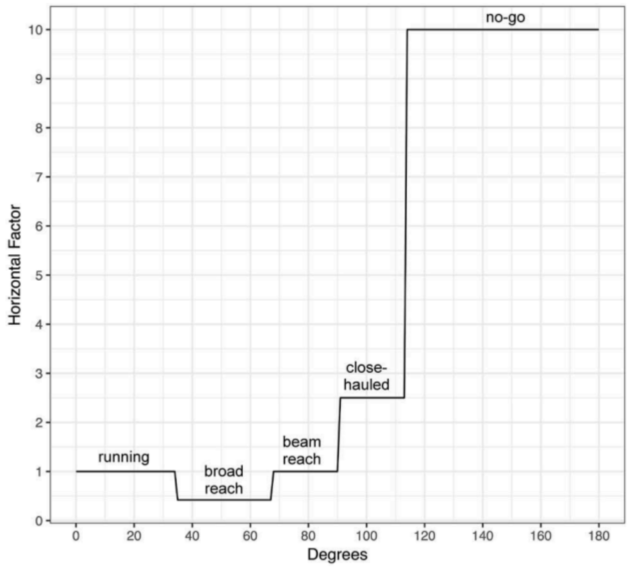
The construction of the data set of sailing times across all coastal segments was carried out by Christopher Barnett, Carolyn Talmadge, Uku-Kaspar Uustalu and the other Tufts Data Lab team. We use the sailing model from Alberti (2018), which computes sailing time given wind speed and wind direction. 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 A3 shows the wind patterns around the Mediterranean.

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.[24] This process allows the cost to be different when travelling 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.

To check our data quality, we compare 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 A2. Visually they are very consistent. Also, regression for minimum (among all seasons) estimated against historical observed minimum voyage times without a constant has a R^2 of 0.82.

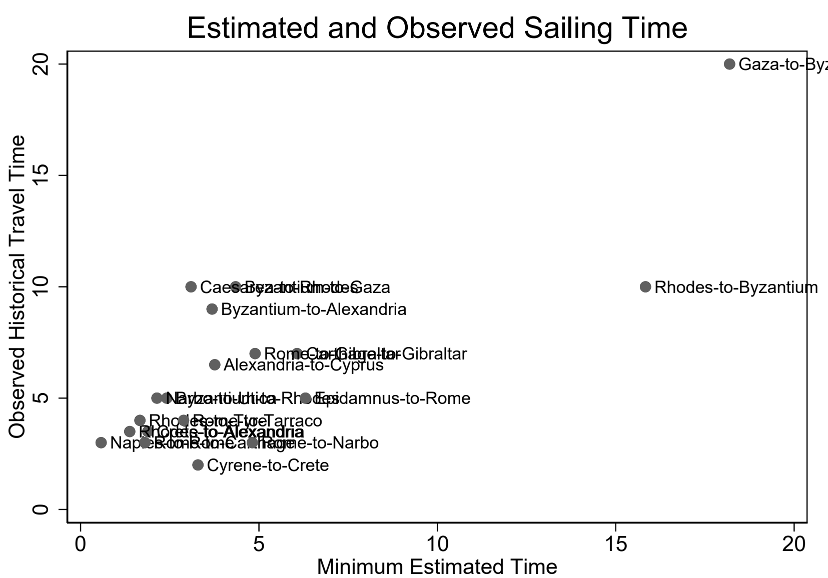
We conduct additional data quality check by cross-checking for a few of our sites which are

Figure A1: Horizontal Factor Table



Source: Figure 3 of Alberti (2018)

Figure A2: Travel Times: Historical Observed vs. Estimated



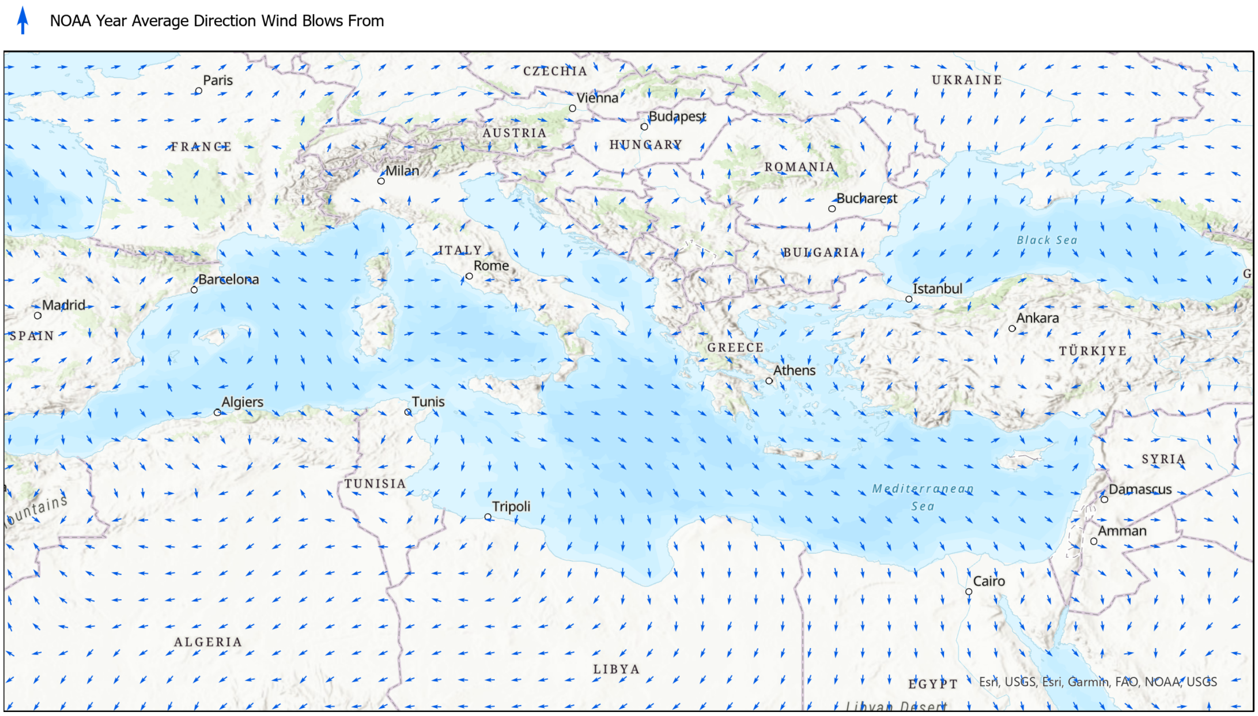


Figure A3: Yearly Average Wind Direction

also available in the Stanford Orbis project (https://orbis.stanford.edu/) for the sailing time in winter. One example is Kranioi and Syrakousai. Figure A4 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 A5 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 A6 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.

In all 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 sources.

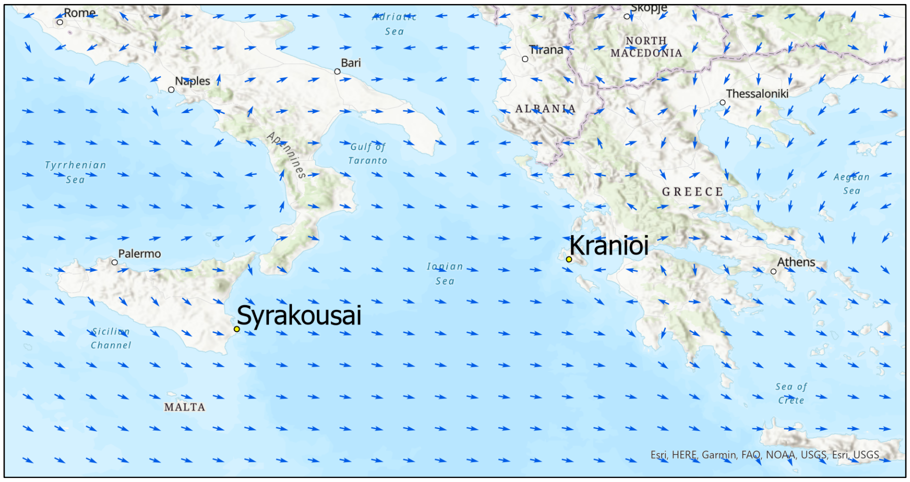


Figure A4: Example 1: Syrakousai and Kranioi

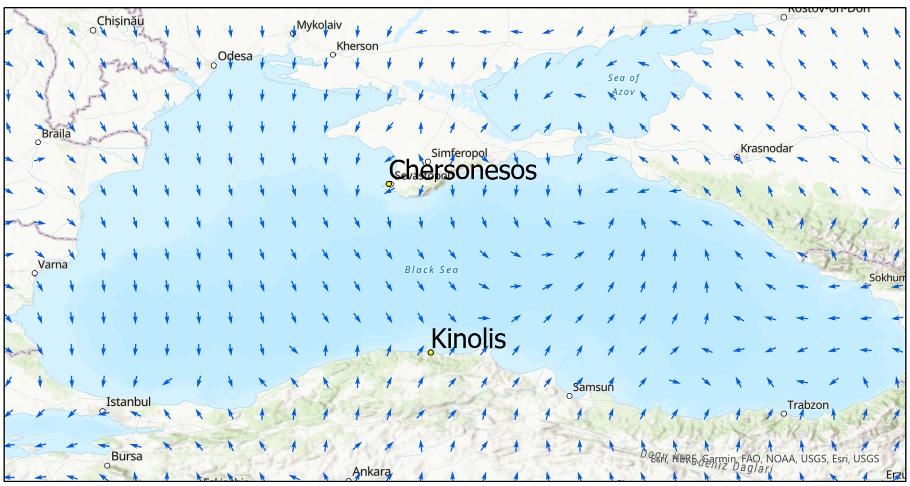


Figure A5: Example 2: Chersonesos and Kinolis

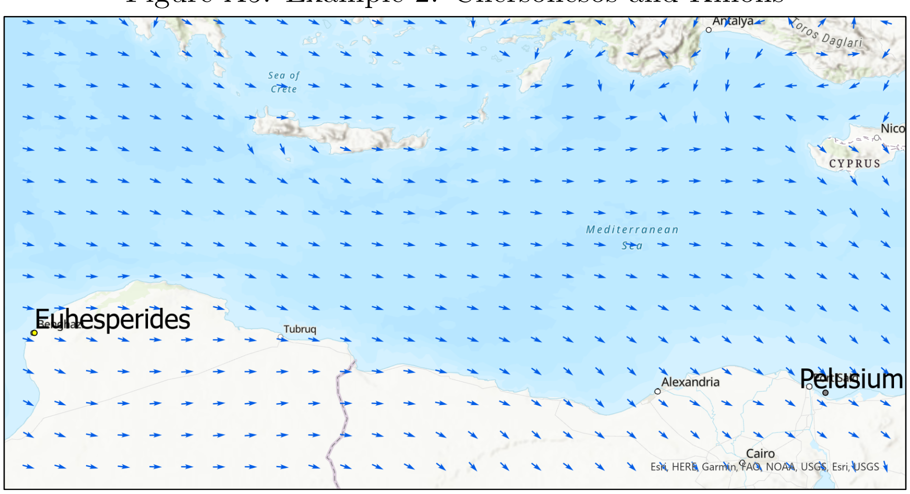


Figure A6: Example 3: Euhesperides and Pelusium

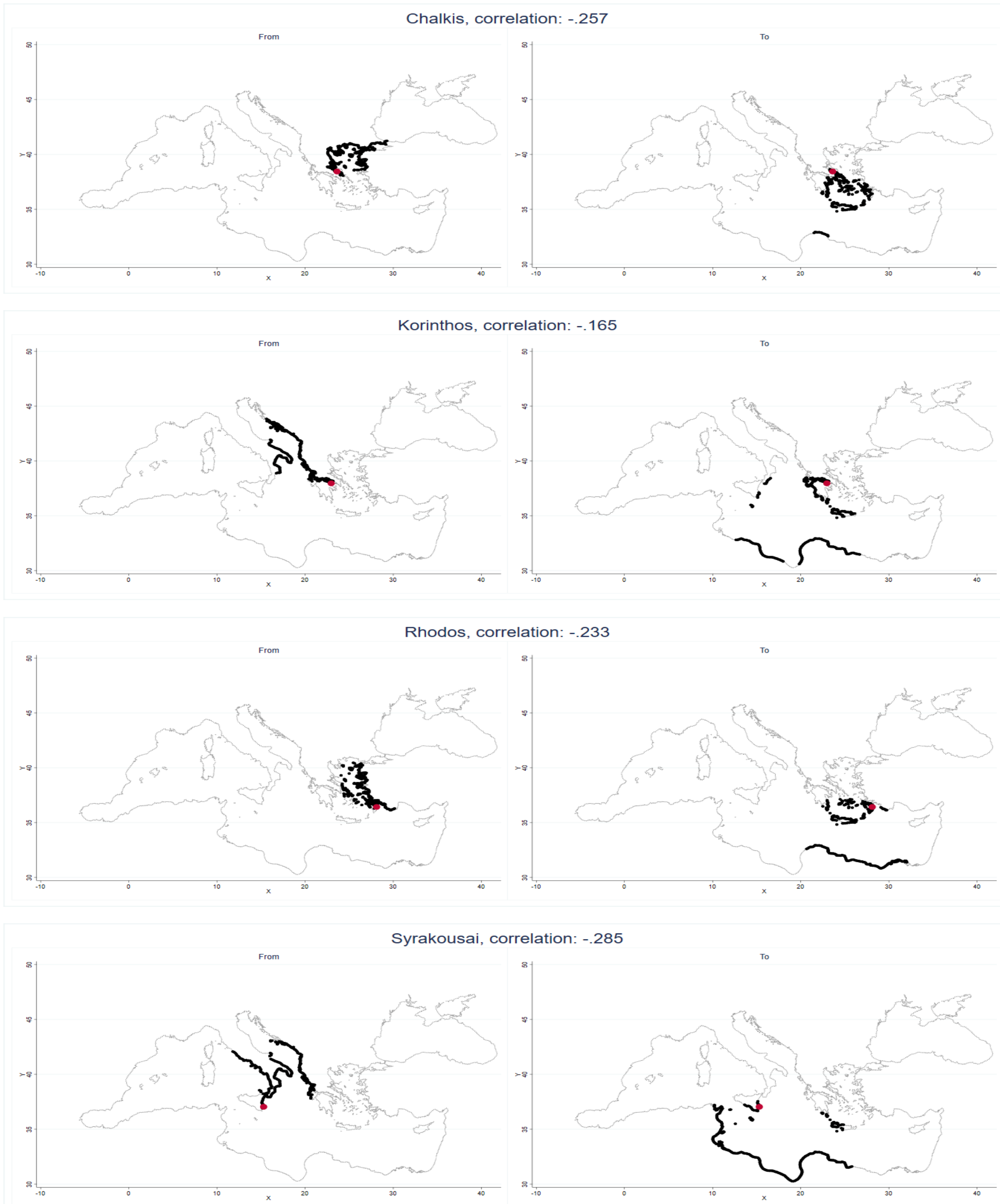
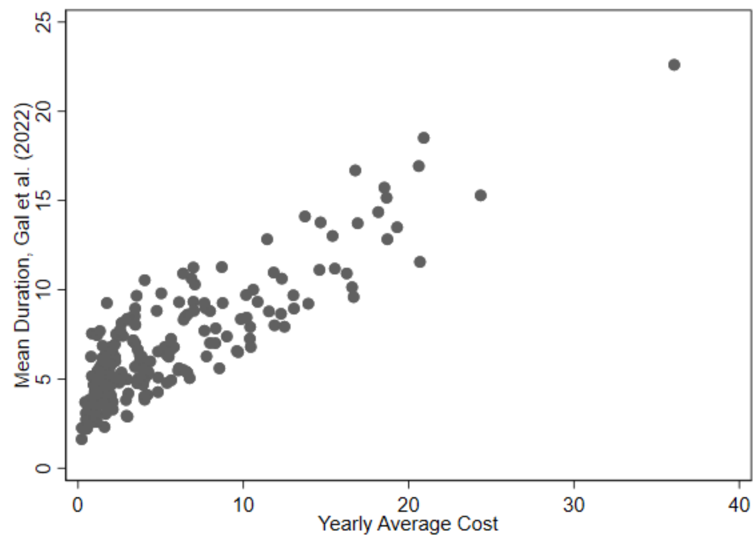


Figure A7: These maps show the 10 percent closest segment points *EMA* (labelled To) and *IMA* (labelled From) for four particular poleis. The plots also show the correlation between the *EMA* and *IMA* market access for segment points.

Figure A8: Travel Times: Gal *et al.* (2022) vs. Own Imputations



Apart from the Stanford Orbis project, 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 A8.

D 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