Asymmetric Trading Costs and Ancient Greek Cities

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

Brown University, Tufts University and University of Oxford

May 13, 2022

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. 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.

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

**Professor of Economics, Tufts University. Yannis.Ioannides@tufts.edu.

***Professor of Economics, University of Oxford. ferdinand.rauch@economics.ox.ac.uk

*We acknowledge helpful comments from a number of scholars without implicating them, especially from Costas Arkolakis, John Kroll and Josiah Ober, but also from Alain Bresson, Yannis Evrigenis, Bruce Hitchner, Misha Kilmer, Hon Ho Kwok, James Murphy, Theodore Papageorgiou, 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 among poleis data. We are also grateful to 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 especially from Gengbin Piao, but also by Joseph Chance, Xu Dong, James Prosser, Zach Sowerby, and Wencong Xu.

**Corresponding author: Yannis M. Ioannides, Department of Economics, Tufts University, Medford, MA 02155, USA. `Yannis.Ioannides@tufts.edu`

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. We then take these predictions to a dataset of the locations of cities in Hellas. 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. In particular, market access seems to matter most when it connects with locations for imports in winter. We also demonstrate new facts regarding locational fundamental factors 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 (2004)].[1] Classicists and historians have documented extraordinary patterns of trade, growth, and political reforms. 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 (2005),[2] and have been recently been digitized. Those together with additional data from other sources are amenable, in principle, to quantitative analysis.

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? Do natural factors matter more than distances to other cities? What is the role of seasonal variation? Are there differences between large and small poleis?

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 Mediterranean and Black Sea coasts and their hinterlands. 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) in his study of the economy and economics of Ancient Greece (with a special emphasis on the economy of Athens) considers the study of the economy of ancient Greece as shedding light on the primitivist vs. modernist and the formalist vs. substantivist controversies. The latter was a debate primarily limited to economic anthropology. 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.

A formalist view analyzes the ancient economy in terms of the basic behavioral assumptions of modern economics, namely, utility maximization and profit maximization. In contrast, a substantivist view might demand different behavioral assumptions, such as status maximization [c.f. Finley(1999)]. Therefore, formalism and modernism are close, yet substantivism and primitivism are not necessarily so. We do not intend to enter this controversy and are prepared to employ modern economics tools without apology.

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.[3] 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. However, how might those changes have affected geographical difference in political economy? See *ibid.*, Fig. 2 and 3.[4] 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.[5] In contrast, in Macedonia trends in the relative presence of cereals and olives goes 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, rather than the number of coastal points, as a more direct measure of market access, by counting the number of sites within a given distance. 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.

Another important consideration is the potential significance of alliances. We have data on which poleis were member of the Delian League,[6] the group of allied poleis under the hegemony of Athens, and of other looser and lesser known associations of poleis, *Koinon*,[7] for which a variable that codes membership in a Federal state before 300 BCE is also available. At the peak of the Athenian maritime empire, the policing of the seas was financed by alliance tributes, which also occasionally generated frictions. After all, conflicts over access to resources is just as much an ancient as a modern concern of states. Indeed, as Jackson and Nei (2015) argue, increased trade decreases countries' incentives to attack each other and increases their incentives to defend each other, leading to stable and peaceful networks of military and trade alliances that is consistent with observed data over 1950–2000. The simplicity of their argument lends itself to application to the ancient era as well. One may wonder about the converse, namely that threat of war by a polis over another (poleis) makes the threatened party acquiesce to expanding trade, and/or joining an alliance. In the classical era, such threats were even "democratically" debated, as witnessed by the Melian Dialogue.

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 such far away places as 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 market access regressions which are motivated by our model. Our results are robust with respect to an alternative measure, that is, eigenvector centralities associated with the shipping network. This model is appropriate because the ancient Greek cities, the Hellenic city-states, were in effect sovereign entities, which shared numerous characteristics including geographic ones. It is particularly noteworthy that the empirical analysis reveals that asymmetries in shipping costs are very important. We think of this as a robust finding. It is notable in our market access regressions that connections *To* each polis from all other poleis are empirically more important than *From* each polis to all other poleis.

Specifically, the market access regressions are segment-based, are conducted with 2SLS and with an outcome variable the count of poleis in the vicinity of each segment point. They typically yield statistically significant effects for market access to poleis in the vicinity of other segments, weighted by the respective shipping costs. Seasonally varying shipping costs are used to produce seasonal market access, while controlling for segment-specific precipitation, temperature, elevation and its standard deviation, malaria, and ruggedness. Market access of each polis is defined in terms of market access to all other poleis in the vicinity of all other segments. Market access variables are also defined by weighting with crop suitability variables, and the directional nature of shipping costs is accounted for by defining shipping costs to and from each other segment. When we do vary market access by the size categories of poleis that can be reached, we find that coefficients tend to increase in the size of polis that a segment point connects with, both in direction *To* and direction *From*. We note again that for poleis of all sizes, connecting *To* them is more valuable than being connected *From* them.

1.3 Outline of the Paper

The remainder of the paper next presents, in section 2, a model of trading poleis (cities) under the Armington assumption using Dixit-Stiglitz preferences, explores determination of wages with fixed populations, and then compares with outcomes under the spatial equilibrium assumption of economic geography when populations and wages are endogenous, and discusses their empirical implications. Section 3 discusses aspects of the data, with more detail on the poleis data set given in section 3.1 and the segments data given in section 3.2. Our empirical results are presented in section 4, where section 4.2 presents our market access regressions and section 4.3 outlines our investigations with the poleis data.

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 Urban System with Dixit-Stiglitz Preferences and Armington Trade

2.1 Model setup

This section develops a model of trade among poleis located in given sites. While the model adheres to the canonical intercity trade framework of Henderson (1974), it abstracts from urbanization and agglomeration effects and relies on the modern intercity trade literature in the style of Allen, Arkolakis and Li (2016), as simplified by Allen and Donaldson (2020). We use this model as an example to show that asymmetric trade costs are a feature of simple and general trade models.

For a given set of sites occupied by poleis, we specify a trading system and use it to motivate two sets of estimations: one, of the determinants of settlements, in other words, where poleis locate; and two, the interdependence of polis sizes, given their existence and location. 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 every site by a large number of producers with the same productivity characteristics:

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

TFP A_i could be specified in more detail, if necessary, in order to express geographical features, soil conditions, etc., which our empirical analysis does allow for. The number of poleis, $i = 1, \dots, J$, is given.

We invoke the Armington assumption [Armington (1969)] that goods are differentiated by origin. A consequence of this assumption is that the model does not imply an agglomeration force associated with the range of goods, which is determined by the number of locations in the Armington setting and thus given. In contrast, it is endogenous with Dixit-Stiglitz preferences, increasing returns to scale in production and free entry of firms, as in settings such as Fujita *et al.* (1999).[11]

We work with representative individuals in each site, who have identical preferences in the form of a CES function of $c_{ji}, j = 1, \dots, J$, the quantities of the goods produced in site j and consumed in site i. The constant elasticity of substitution $\sigma > 1$ is the same across all differentiated goods that are produced in the system of poleis, $j = 1, \dots, J$:

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

where u_i denotes site $i-$specific amenities. 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 equilibrium 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, defined, in (5) below, in terms of all bilateral prices, that is

$$p_{ij} = p_i \tau_{ij},$$

where $\tau_{ij} > 1$, bilateral iceberg shipping costs, denotes the units of the good produced at i that must be shipped in order for one unit to arrive at site j. In our empirical applications it is proxied by sailing times which, due to wind patterns, are directional, implying that costs are asymmetric: $\tau_{ij} \neq \tau_{ji}$. Asymmetry turns to be empirically very important.

By expressing local prices in terms of the respective wage rates, the bilateral prices, the prices that consumers at site j pay for a unit of good i, are given by

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

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

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

The price index may be interpreted as *consumer, or inward, market access*; see section 2.4.3 below for further discussion.

2.2 Determination of Wages with Given Populations

The Armington assumption, that is goods being differentiated by their origin, together with utility maximization yield gravity-type trade flows. This is a popular assumption in trade theory and modern spatial economics.[12] Labor supply is inelastic in every polis i. Polis i's sales to j, typically referred to as trade flows from i to j, are given by:

$$X_{ij} = \left(\frac{\tau_{ij}}{A_i u_j}\right)^{1-\sigma} w_i^{1-\sigma} \Omega_j^{1-\sigma} w_j^{\sigma} L_j. \tag{6}$$

We assume first that there is no free movement of labor (this assumption will be relaxed later) and that trade clears at every polis. In such a case, the general economic equilibrium determines nominal wages, but in effect real wages because a normalization is readily available: the X_{ij}'s are homogeneous of degree 1 in all wages. Consequently, real utility in every polis is determined.

For trade equilibrium, first, labor incomes, that is payments to labor in every city equal total sales, including local ones:

$$w_i L_i = \sum_j X_{ij}. \tag{7}$$

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:

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

These conditions respectively become:

$$w_i L_i = A_i^{\sigma-1} \sum_j \left(\frac{\tau_{ij}}{u_j}\right)^{1-\sigma} w_i^{1-\sigma} \Omega_j^{1-\sigma} w_j^{\sigma} L_j; \tag{8}$$

$$w_i L_i = u_i^{\sigma-1} \sum_j \left(\frac{\tau_{ji}}{A_j}\right)^{1-\sigma} w_j^{1-\sigma} \Omega_i^{1-\sigma} w_i^{\sigma} L_i. \tag{9}$$

This is a system of $2 \times J$ equations in the unknowns nominal wages and real utilities for each location, $\{w_i, \Omega_i\}_{i=1}^{J}$. We proceed to characterize the determination of wages and utilities.

Let us define the auxiliary variable

$$\tilde{x}_i := w_i^{\sigma} L_i. \tag{10}$$

By solving (9) with respect to $\Omega_i^{1-\sigma}$ and substituting into (8), we obtain a system of nonlinear equations in the $\tilde{x}_i$'s. That is, from (9) we have:

$$\Omega_j^{\sigma-1} = u_j^{\sigma-1} \sum_k \left(\frac{\tau_{kj}}{A_k}\right)^{1-\sigma} w_k^{1-\sigma} w_j^{\sigma-1}. \tag{11}$$

By rewriting (8),

$$\tilde{x}_i = A_i^{\sigma-1} \sum_j \left(\frac{\tau_{ij}}{u_j}\right)^{1-\sigma} w_i^{1-\sigma} \Omega_j^{1-\sigma} \tilde{x}_j, \tag{12}$$

and substituting for $\Omega_j^{1-\sigma}$ from (11) back into (12) a system of J equations follows which may be written in terms of the $\tilde{x}_i$'s as the unknowns. To accomplish that we rewrite the terms $w_j^{\sigma-1}$ as $w_j^{\sigma-1} \equiv \left(w_j^{\sigma}\right)^{\frac{\sigma-1}{\sigma}}$ and also multiply and divide the respective terms by $L_j^{\frac{\sigma-1}{\sigma}}$, which are exogenous. We thus have:

$$\tilde{x}_i = \sum_j \frac{\left(\frac{\tau_{ij}}{A_i}\right)^{1-\sigma}}{\sum_k \left(\frac{\tau_{kj}}{A_k}\right)^{1-\sigma} \left(\frac{L_k}{L_j}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{\tilde{x}_j}{\tilde{x}_k}\right)^{\frac{\sigma-1}{\sigma}}} \tilde{x}_j,\ i = 1, \ldots, J. \tag{13}$$

This is a homogeneous system of nonlinear equations with the $\tilde{x}_i$ as the unknowns. Following Allen and Arkolakis (2016) and Allen *et al.* (2019),[13] one could work from first principles in order to establish existence and other properties of the system of poleis in trade equilibrium, given populations. Once wages are determined, the price index and real utilities follow.

We explore next another aspect of the model. Specifically, simplifying (9) and solving for Ω_i and writing it instead in terms of Ω_j to avoid confusion, yields:

$$\Omega_j^{1-\sigma} = u_j^{1-\sigma} w_j^{1-\sigma} \left[\sum_k \left(\frac{\tau_{kj}}{A_k}\right)^{1-\sigma} w_k^{1-\sigma}\right]^{-1}. \tag{14}$$

By substituting into (8) and rewriting we have:

$$w_i L_i = \sum_j \frac{w_i^{1-\sigma} \left(\frac{\tau_{ij}}{A_i}\right)^{1-\sigma}}{\sum_k \left(\frac{\tau_{kj}}{A_k}\right)^{1-\sigma} w_k^{1-\sigma}} w_j L_j. \tag{15}$$

Given wage rates, Equation (15) form a linear system of equations in the nominal incomes $w_i L_i$ of the urban system, with the matrix in its right-hand side being column-stochastic, its column sums are equal to 1. Its unique non-zero solution is the right eigenvector corresponding to the system matrix, denoted by $\mathbf{A}$, with its maximal eigenvalue being equal to 1, by the Perron-Frobenius theorem. The corresponding right eigenvector, whose components are the nominal incomes, has positive elements. Since this is defined up to scale, it may be the basis for studying the size distribution of poleis, but it is not pursued further here.

2.3 Determination of Wages and Populations under Spatial Equilibrium

If individuals were allowed to relocate, that is, labor supply were infinitely elastic at every site, then at the spatial equilibrium, utilities would be equalized, 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 simply 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 to be specified shortly.

Spatial equilibrium is attractive because it yields testable predictions: wage rates, populations and utilities are all specific functions of σ, site amenities and productivities, and bilateral shipping costs. Imposing spatial equilibrium might sound as an extreme, but yet is quite plausible an assumption in the long run, given the fact 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 order to pursue opportunities and geopolitical advantages.

Specifically, recall the auxiliary variable $\tilde{x}$ defined in (10) and define, in addition, the auxiliary variable $\tilde{y}_i$ as follows:

$$\tilde{y}_i := w_i^{1-\sigma}. \tag{16}$$

With definitions (10) and (16), and the additional definition

$$\tilde{A}_{ij} := u_j^{\sigma-1} \left(\frac{\tau_{ij}}{A_i} \right)^{1-\sigma}, \tag{17}$$

equations (8), after dividing through by $w_i^{1-\sigma}$, and (9), after dividing through by $w_i^\sigma L_i$, respectively, become:

$$\tilde{x}_i = \sum_j \Omega_j^{1-\sigma} \tilde{A}_{ij} \tilde{x}_j; \tag{18}$$

$$\tilde{y}_i = \Omega_i^{1-\sigma} \sum_j \tilde{A}_{ji} \tilde{y}_j. \tag{19}$$

If utilities are equalized across all poleis, $\Omega_i^{1-\sigma} = \bar{\Omega}$, the Rosen-Roback or economic geography case prevails, the terms $\Omega_i^{1-\sigma}$ factor out of the summation in (18) and the above conditions may be rewritten as:

$$\tilde{\mathbf{x}} = \bar{\Omega} \tilde{\mathbf{A}} \tilde{\mathbf{x}}; \tag{20}$$

$$\tilde{\mathbf{y}} = \bar{\Omega} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}. \tag{21}$$

Equations (20) and (21) viewed as system of linear homogeneous equations in $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ yield that $\bar{\Omega}$ is the Perron-Frobenius eigenvalue associated with the kernels of (20) and (21), respectively, which is common among them, because $\tilde{\mathbf{A}}^T$, the kernel of (21), is the transpose of $\tilde{\mathbf{A}}$, the kernel of (20). The nonzero solutions for $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}$ denote the Perron-Frobenius right and left eigenvectors of $\tilde{\mathbf{A}}$, the kernel of (20) and (21). These solutions uniquely determine equilibrium real wages and populations up to scale as the left and right Perron-Frobenius eigenvectors of $\tilde{\mathbf{A}}$. Note that the entries $\tilde{A}_{ij}$ combine the shipping costs from i to j with TFP at i and local amenities at j.[14]

Finally, regarding the equilibrium wage rate, we may work as follows. The definitions of the auxiliary variables $\tilde{x}_i, \tilde{y}_i$ readily imply that:

$$\tilde{x}_i \tilde{y}_i = w_i L_i. \tag{22}$$

Therefore, it is more convenient for us to normalize $\tilde{\mathbf{x}}$ by $\sum_i \tilde{x}_i = Y$, so that $\sum_i \tilde{x}_i \tilde{y}_i = \sum w_i L_i = Y$, which then denotes nominal income. Real wages follow by normalization. Equilibrium populations are obtained by solving for L_i in terms of $\tilde{x}_i, \tilde{y}_i$ and scaling so that $\sum_i L_i = \bar{L}$., where total population $\bar{L}$ is an exogenous variable. Using the fact that the elements of $\tilde{x}_i, \tilde{y}_i$ are known (up to scale), we may solve for the populations as follows. That is, from

$$\frac{w_i^\sigma L_i}{w_j^\sigma L_j} = \frac{\tilde{x}_i}{\tilde{x}_j}, \ \frac{w_i^{1-\sigma}}{w_j^{1-\sigma}} = \frac{\tilde{y}_i}{\tilde{y}_j},$$

which implies:

$$\frac{L_i}{L_j} = \frac{\tilde{x}_i}{\tilde{x}_j} \left(\frac{\tilde{y}_i}{\tilde{y}_j} \right)^{\frac{\sigma}{\sigma-1}}. \tag{23}$$

This along with $\sum_i L_i = \bar{L}$ determine all populations uniquely. That is:

$$L_i = \bar{L} \frac{\tilde{x}_i (\tilde{y}_i)^{\frac{\sigma}{\sigma-1}}}{\sum_j \tilde{x}_j (\tilde{y}_j)^{\frac{\sigma}{\sigma-1}}}. \tag{24}$$

Similarly, relative wages and relative incomes are given in terms of $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}$:

$$\frac{w_i}{w_j} = \left(\frac{\tilde{y}_i}{\tilde{y}_j} \right)^{\frac{1}{\sigma}}; \frac{w_i L_i}{w_j L_j} = \frac{\tilde{x}_i}{\tilde{x}_j} \frac{\tilde{y}_i}{\tilde{y}_j}, i, j = 1, \ldots, J. \tag{25}$$

Note that (23), (24), and (25) are independent of the factors of proportionality.

In sum, the assumption of spatial equilibrium is attractive because the equilibrium solutions rest entirely on exogenous variables. It is commonly invoked in studies with modern data. Equalized utility may be a realistic assumption in our setting as well, particularly given that we study outcomes over a long period of time in a setting with inter-city migration.[15] In such a case, the elements of $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}$ are determined up to a factor of proportionality, in other words, the ratios are determined, and so are the relative incomes. Therefore, the terms $\frac{\tilde{x}_i}{\tilde{x}_j} \frac{\tilde{y}_i}{\tilde{y}_j}$ may be thought of as relative centralities.

Regarding implementation, working with imputed traveling times as proxies for the τ_{ij}s is straightforward. Working with the amenities and productivities u_i, A_i is less so. That is, not only they are multidimensional, but from (17), the terms depend on the productivity of the origin, A_i, and the consumer amenities at the destination, u_j. Note that if we normalize the eigenvectors $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}$ so that $\tilde{\mathbf{x}} \cdot \tilde{\mathbf{y}} = 1$, which we are allowed to do, then we may solve for the relative real wages $\frac{w_i}{w_j}$, the relative nominal incomes $w_i L_i$ and the relative populations in terms of the components of the eigenvectors. With a given total population, actual populations may be solved for. The definitions (10) and (16) of the auxiliary variables $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$, readily yield that the vector of the nominal incomes is given by the Hadamard product, that is the element-wise product of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$, which may unfortunately not be expressed analytically in terms of the entries of $\tilde{\mathbf{A}}$.

Whether we work with the values of sites or notional incomes, in reduced-form settings, we can regress a number of proxies for urban settlements at each segment s against observable characteristics of all other segments, $s \in \mathcal{S}, \neq i$. An intuitive measure of urbanization is the number of observed poleis within different radii of each segment in the data. Working in the spirit of the above model in the special case of $\sigma = 2$ leads to the special case of the Harris market potential [Harris (1954)] as a proxy for the market access accorded to each segment. Wage rates are endogenous and not available in closed-form, but we account for this endogeneity in our estimations. As we clarify in section 2.4.3 below, the conditions for spatial equilibrium allows to obtain a prediction of the relative importance of producer vs. consumer market access, that is, *From* vs. *To* market access, respectively.

We report results along these lines in section 4.2. We take up empirically the question of how seasonal fluctuations in shipping costs might affect settlement patterns and polis sizes 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.

2.4 Value of Sites under Dixit-Stiglitz Preferences

Let $s \in \mathcal{S} := \{1, \ldots, S\}$ denote all sites that are feasible for urban development.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 travel times across all segments and a wealth of local crop suitability and climatic data, all included in our Segments Data. In addition, the segments may be easily linked with a data set pertaining to locations and characteristics of Hellenic poleis, to be referred to as the Poleis Data, which originate in Stanford Polis Project, with additional data collected and coded by us.

Recall (2) and assume that activities compete across sites, so that a particular set of productivity characteristics and patterns of shipping costs yield maximum indirect utility $\mathcal{V}_i$. Taking logs, the value of site s is given by:

$$\mathcal{V}_s = \ln \Omega_s = \ln w_s + \frac{1}{\sigma - 1} \ln \left[\sum_k \left(\tau_{ks} \frac{w_k}{A_k} \right)^{1-\sigma} \right] + \tilde{u}_s, \tag{26}$$

where $\tilde{u}_s$ is assumed to be a stochastic counterpart of $\ln u_s$.

Two aspects of our data on shipping costs have potential consequences for the sizes of urban settlements and therefore for the values of sites. The first is the variability of shipping costs during the year. Our data include information on their seasonal variation across segments.[16] We examine further below the consequences of seasonal variations of shipping costs. The second is the fact that, as we show in section 2.4.3 below, asymmetric shipping costs drive a wedge between producer and consumer access, which unlike in the symmetric case under Dixit-Stiglitz preferences, they are no longer proportional to one another.

As Redding and Rossi-Hansberg (2017) conjecture, asymmetric shipping costs have implications both for the characterization of equilibrium and for patterns of trade and income [c.f. Waugh (2010), Allen *et al.* (2019)]. In fact, such asymmetries are not specific to ancient maritime transportation. They are very much present in modern shipping and do vary by the type of cargo, such as containers versus bulk versus oil.[17] Such asymmetries emanate from the directionality of ocean currents and wind patterns but also the patterns of shipments across different markets. Our market access regressions, reported in section 4.2, allow us to establish that asymmetry in shipping costs is empirically significant for the patterns of settlements.

2.4.1 Value of Sites with Seasonal Variations in Shipping Costs and Given Populations

In the case of equilibrium with given populations, the value of sites are given by evaluating (26) at the solutions for wages implied by the values of $\tilde{x}_i = w_i^\sigma L_i$'s that solve (13). Naturally, they do depend on all parameters and, in addition, on the given populations, the L_i's. We may interpret the solutions for the $\tilde{x}_i$'s as notional incomes.

Implicit in this exposition is that shipping costs and productivities are constant. However, as we discuss in sections 3.1 and 3.2 below, shipping costs exhibit seasonal variations, and in fact their seasonal means are included in our data. As shipping costs vary seasonally, so do trades. Since populations are given, wages vary seasonally. Therefore, the resulting variation in wages makes site values (26) ex ante uncertain. Consequently, sites may be evaluated in terms of the expected values of $\mathcal{V}_s$ with respect to uncertain shipping costs. That is, assuming that shipping costs become known at the beginning of every season, then trades are executed and therefore realized utilities will vary seasonally. Of course, depending on the nature of the goods, storable goods may be inventoried and thus handled differently from perishable ones. However, we lack any detailed information on the nature of trades and thus restrict attention to general predictions.

The expected value of a site s, on an annual basis with given seasonal fluctuations, $\mathcal{T}_{\text{season}}$, is given by (26), evaluated at the respective values of wages, $w_i(\mathcal{T}) = \left(\frac{\tilde{x}_i(\mathcal{T})}{L_i}\right)^{\frac{1}{\sigma}}$, which is given implicitly by the solutions of (13). That is,

$$\mathcal{E}[\mathcal{V}_s] = \frac{1}{4}\sum_s \mathcal{V}_s[\mathcal{T}_{\text{season}}]. \tag{27}$$

As is clear from (26), site values are convex decreasing functions of shipping costs, τ_{ks}. However, their net impact also depends on the impact of seasonal fluctuations on wages. Since the site values are homogeneous of degree 0 in all wages, which follows from (3) and (5), it is the real impact of fluctuations on site values that matters. We may see this by characterizing the impact of fluctuations on relative wages and on real incomes at all sites.

Real wages may be obtained by choosing one the wages as numeraire, say $w_1 = 1$, and solving for the $\tilde{x}_i$s accordingly. Real incomes are then given by $w_i^N L_i$. Unfortunately, the analytics are too complicated to allow us to obtain theoretical predictions about how real wages $w_i^N L_i$ vary with the variations in the τ_{ij}s. Additional assumptions regarding time preferences, durability and the cost of storage of goods may be required. It is thus possible in principle that fluctuations in shipping costs improve real outcomes. If shipments vary in perishability, the possibility of storage generates a value of the option from having variable shipping costs, as one can take advantage of the periods of low costs. Such an option value may enhance the value of certain sites.[18]

2.4.2 Value of Sites with Seasonal Fluctuations in Shipping Costs and Populations Determined via Spatial Equilibrium

If populations are not given and instead are determined by individuals' reallocating so as to be indifferent among all sites, then the counterpart of the conditions obtained in section 2.3 above must be adapted for the case of seasonal variation of shipping costs. Relocation of population is of course meaningless unless seasonal migration is allowed, which is not an unreasonable assumption especially across neighboring poleis. The conceptual counterpart of spatial equilibrium that follows equalization of utility across all sites would be to define populations that would bring about equalization of expected utility. However, this implies a conceptually very different setting. In the interest of a better conceptual understanding, we may approach the problem as follows.

Consider outcomes for site s. In a Nash equilibrium setting, conditional on populations located at sites $\{L_j : j \neq s, \in \mathcal{S}\}$, it follows that $L_s = \bar{L} - \sum_{j \neq s} L_j$, and wages are implied for each realization of shipping costs, as in section 2.4.1. Evaluating expected utilities as in (27) above, and setting them equal across all sites $s \in \mathcal{S}$ yields $J - 1$ equations. An additional equation is given by the condition $\sum_s L_s = \bar{L}$, and therefore the J populations of all sites are obtained as deterministic values.[19] Such fluctuations penalize shipping to and from sites with higher costs, thus reducing their attraction as trading partners and therefore wages, given populations. However, since prices are proportional to the product of wages and shipping costs, the impact on real utility and on the pattern of settlements at all sites is ambiguous.[20]

While this is conceptually close to spatial equilibrium in the static case, developed in section 2.3 above, it is analytically very unwieldy. With this in mind, we propose next an alternative conceptualization that utilizes the derivations obtained in section 2.3. That is, by using (25) to write $w_k = w_s(\tilde{y}_k/\tilde{y}_s)^{\frac{1}{\sigma}}$, conditional on the realization of shipping costs, the expression for $\mathcal{V}_s$ in the case of spatial equilibrium becomes:

$$\mathcal{V}_s = \ln \Omega_s = \frac{1}{\sigma - 1} \ln \left[\sum_k \left(\frac{\tau_{ks}}{A_k} \right)^{1-\sigma} \left(\frac{\tilde{y}_k}{\tilde{y}_s} \right)^{1-\sigma} \right] + \tilde{u}_s. \tag{28}$$

It is therefore expressed in terms of the shipping costs and productivities, and of the entries of the left eigenvector of $\tilde{\mathbf{A}}$, themselves functions of exogenous variables. Although these expressions are convex decreasing in shipping costs as direct arguments,[21] they also depend on shipping costs via the $\tilde{y}_i$'s, but we are unable to be more specific about the latter effects. Since utilities are equalized under each realization, the expected utilities according to (27) are also equalized across sites. The associated populations are given by (24) for each realization.

The two alternative ways that we propose to conceptualize spatial equilibrium under seasonal variation of shipping costs are not equivalent in terms of underlying assumptions. While both involve equalization of utilities, the former is premised on the notion that populations are determined prior to the realization of shipping costs, and it is ex ante expected utilities that are equalized; and the latter on the notion that activities reallocate seasonally so as to equalize utility. If it might be reasonable to assume that population relocate seasonally, then the latter is a simpler formulation, even though it does not lead to closed-form solutions. Population movements were not unknown in the classical Hellenic era,[22] as witnessed by colonization, but in view of residence restrictions by several ancient poleis, it would be unreasonable to assume seasonal relocations. At any rate, it is reasonable to assume that the ex ante values of sites depend on the dispersion characteristics of the underlying seasonal fluctuations of shipping costs, and not just their means, a notion which we actually do examine empirically.

2.4.3 Linking to Market Access

Motivated by Paul Krugman's[23] linking of modern economic geography theory with the Harris (1954) market access (also known as market potential) concept favored by geographers, several authors have relied on its modernization [Redding and Venables (2004)] and used it extensively to explore urban evolution and urbanization [*c.f.* Michaels and Rauch (2018); Bakker *et al.* (2021)]. Anticipating the empirical analysis in section 4, we provide conceptual links of the market access approach with our theory, which is based on Dixit-Stiglitz preferences via Allen and Arkolakis, *op. cit.* Notably, the theory distinguishes producer and consumer market access, as we see shortly, and leads to an expression for population size L_i in terms of both market access concepts.

A key feature of our data on shipping costs is that they are asymmetric. Among the few papers in the literature that emphasize asymmetries in shipping costs, Waugh (2010) first established empirically the fact that empirical model-implied costs systematically deviate from symmetry, in the sense that they are not the same function of distance, shared border, language, colonial relationship etc. and co-vary with the level of development. Poor countries face higher export costs relative to rich countries. Behrens and Picard (2011) explore asymmetric costs further in a study of trade and economic geography by relying on quasi-linear quadratic model with two regions.[24] Asymmetric shipping costs is also a feature of treatments of modern shipping, as well; see, in particular, Brancaccio *et al.* (2020).

We develop next an expression for sizes of settlements as functions of producer and consumer market access in the general case of asymmetric trading costs. Unlike in the symmetric case, where as Redding (2020) proves they are proportional, such asymmetries have profound effects on the relationship between producer and consumer market access.

Producer market access, or *outward trade market access*, may be defined as [Redding and Rossi-Hansberg (2017)]:

$$\mathrm{PMA}_i := \sum_j \tau_{ij}^{1-\sigma} \frac{w_j L_j}{P_j^{1-\sigma}}. \tag{29}$$

Correspondingly, consumer market access, or *inward trade market access*, P_i, is defined as a transformation of the price index. Specifically, following Redding and Rossi-Hansberg (2017), consumer market access may be defined in terms of the same scale as producer market access:

$$\text{CMA}_i := P_i^{1-\sigma} = \sum_k \left(\tau_{ki} \frac{w_k}{A_k} \right)^{1-\sigma}. \tag{30}$$

It thus reflects purchasing power from the viewpoint of site i, whereas producer market access corresponds to the demand for good i by all other sites, after accounting for shipping costs, incomes and the respective price indices.

Next, we explore properties of the equilibrium utility with respect to changes in τ_{ij} vs. τ_{ki}, intuitively the shipping costs *from* site i to the other sites, vs. the shipping costs *to* site i from all other sites. Exposition is simplified if we recognize that we may rewrite (15) as

$$w_i L_i = \left(\frac{A_i}{w_i} \right)^{\sigma-1} \text{PMA}_i. \tag{31}$$

The population at every site i may be expressed in terms of the consumer and producer market access. From the spatial equilibrium condition (20) for $w_i^\sigma L_i$, we have:

$$w_i^\sigma L_i = \Omega_i^{1-\sigma} \sum_j \tilde{A}_{ij} w_j^\sigma L_j.$$

From the equalization of $\Omega_i^{1-\sigma} = \frac{w_i^{1-\sigma}}{P_i^{1-\sigma}}$ across all sited, we may bring it within the summation and get:

$$L_i = A_i^{\sigma-1} w_i^{-\sigma} \text{PMA}_i = u_i^{-\sigma} A_i^{\sigma-1} P_i^{-\sigma} \text{PMA}_i. \tag{32}$$

Therefore,

$$L_i = u_i^{-\sigma} A_i^{\sigma-1} \text{CMA}_i{}^{\frac{\sigma}{\sigma-1}} \text{PMA}_i. \tag{33}$$

We note that the contribution to L_i of CMA_i, which by definition involves the *to*-polis i shipping costs, the τ_{ki}s, and is raised to the power of $\frac{\sigma}{\sigma-1} > 1$. In other words, it is amplified relative to the contribution of PMA_i, which involves the *from*-polis i shipping costs, the τ_{ij}s, are weighted, that is normalized, by their sums, thus weakening their contribution. Anticipating our empirical results, we note that the empirical counterparts of the market access variables, when both *to*-polis and *from*-polis covariates are included, confirm these predictions.[25]

2.5 Summary of Theoretical Predictions

We have so far employed a model of trade among independent polities, poleis, based on the Armington assumption, goods exchange and absence of monetary transactions, and explored it in two alternative settings. The model in section 2 of an urban system with Dixit-Stiglitz Preferences and Armington Trade lends itself to the following uses and associated predictions. An advantage of the model is that it predicts functional relationships of the market access type.

1. With free trade and fixed populations, the L_is, the interpolis model equilibrates the economy, given local productivities and geography. The predictions are of the gravity type. The system of J equations (13) is homogeneous in the variables $\tilde{x} := w_i^\sigma L_i$. Therefore, $J-1$ of them are independent and determine relative wages.

2. With free trade and free labor mobility, imposing spatial equilibrium determines wages and populations. This is presented in section 2.3, where the model is solved via the two auxiliary variables $\tilde{x}_i := w_i^\sigma L_i, \tilde{y}_i := w_i^{1-\sigma}$. As the right and left eigenvectors of $\tilde{\mathbf{A}}$, respectively, they are determined uniquely up to scale, are strictly positive and therefore imply that all sites are occupied. The formalization of this result is due to Allen and Arkolakis (2014) in a very general geographical setting but with symmetric trading costs. They have extended it to quasi-asymmetry and to presence of spillovers. Quasi-symmetry refers to τ_{ij} as having a factorial structure, with a symmetric core and two origin- and destination-specific factors.

   So, imposing spatial equilibrium along with free trade determines economic activity everywhere in the system. Relative wages are uniquely determined — taking the ratios w_i/w_j cancels the factor of proportionality, a total population of $\bar{L} = \sum_i L_i$, is uniquely allocated to all poleis, and real income in each polis is given by $\frac{w_i}{w_{num}} L_i$, where $\tilde{w}_{num}$ denotes the numeraire wage. These endogenous quantities are determined for arbitrary patterns of productivities and asymmetric trading costs and have very specific functional dependence on the underlying parameters of the entire geography, not just locally, albeit implicitly, as we do not have them in closed form.

3. An important implication of the shipping costs asymmetry follows directly from (20) and (21): $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ correspond, respectively, to shipping costs *To*, and shipping costs *From* poleis. This facilitates the interpretation of our empirical results, as one of the endogenous variables we work with, $polis_s$, by the product of these auxiliary variables according to (22). The intuition for the asymmetry is that shiping *From* relates to exports and hence nominal income, while *To* influences local prices.

4. To summarize, we highlight the following:

   (a) In the absence of trade, only the locally produced good is consumed, and the utility functions have only one argument, the locally produced good. Therefore, trade is obviously advantageous because of love of variety. Similarly, increase in the number of cities by urban expansion to different sites is in principle desirable, if it is costless. City sizes, proxied by the L_i's or even the $w_i L_i$'s, will be independent of the $A_j, j \neq i$, the productivities at other locations, but a function of own productivity only. This implies a straightforward identifiable restriction. However, climatic conditions are not very different, and so identifying local geographic characteristics that are independent across locations would be important. Therefore, reduced forms that regress proxies for urban settlements or polis sizes against productivities everywhere and shipping costs to and from everywhere compared to regressions with own characteristics alone could serve as evidence of trade.

   (b) The trade model of section 2 takes the L_i's as given, and determines equilibrium wages and real utilities, and therefore real incomes as well at every site (poleis) of the system. Under the assumption real wages are given (possibly determined at the subsistence level at each site), then trade determines relative sizes, the L_i's up to a factor of proportionality.

   (c) We also examine in the light of the data a particular prediction of the model, discussed in section 2.4.3. That is, under spatial equilibrium, settlement sizes obey (33). This implies that consumer market access, CMA_i, the *To,* is quantitatively more important than producer market access, PMA_i, the *From.*

3 Data

The present study utilizes two principal data sets, suitably augmented; one, the Poleis Data Set; and two, the Coastal Segments Data Set. We discuss them in more detail next.

3.1 The Poleis Data Set

The Poleis Data Set originates in the *Inventory of Archaic and Classic 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*. Some of the data were digitized by Josiah Ober and his team of Stanford and Oxford scholars [http://polis.stanford.edu/]. From among the available variables of particular interest are the following: Size, the size of the polis territory in the archaeological record, with values ranging from 1 to 5 (with 0 coding absence of evidence of size), region (namely historical region), walls (evidence of fortification), Delian (participation in the Delian League), and *Koinon* (participation in a federation). We supplement these data with local micro-geographic information on ruggedness, distance to the nearest harbors and the sea, crop suitability and temperature data. We have verified and used the GPS coordinates in the data to merge them with the micro-geographic and spatial information describing the environment surrounding each Hellenic polis.[26]

Central to this inquiry is understanding how the Hellenic system of cities has thrived on both intercity and international 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 [Ober (2015)]. 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.

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.

In augmenting the data from the *Inventory* and in addition to the geological descriptions and agricultural suitability of the immediate environment of each polis, we add distances among sites.

3.2 The Coastal Segments Data Set

An additional set of imputations and data collection involve 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. This segments dataset can be seen as a regular grid of points along the coastline in the spirit of Michaels and Rauch (2019) and Bakker *et al.* (2021). Placing points on the coastline ensures that we do not need to define inland transportation costs when computing pairwise transport costs between segment points. Our goal is to create an even 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 could provide versions in which we use different regression weights using our sample of gridpoints to show robustness to alternative gridpoint sampling methods. We do not show such robustness in this version of the paper because we are not clear that an alternative sampling method is preferable, and we do not expect it to make a meaningful difference in our regressions. 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 7.1. Figure 1 compares imputed with some historical travel times.

3.3 Descriptive Statistics

Table 1 provides summary statistics of the key variables in the empirical exercise, which are the variable $polis$, our main outcome variable that consists of a simple count of poleis that are within $10km$ of a segment point (similarly to Bakker *et al.* 2021). The table shows the mean of 0.406 with the standard deviation of 0.807 in brackets. The number of observations for each of these means and standard deviations is the number of total segment points, which is 2,737.

Next we report descriptive statistics for the market access variables, separately for each season and the year as a whole and separately for market access *To* and *From* other segment points. These are variables that sum inverse distances to all other segment points. Here the means for *To* and *From* are identical for each season by construction, since they aggregate the same matrix but once by columns and once by rows. Standard deviations and regressions

$polis_s$	0.406				
	(0.807)				
Market access					
	Spring	Summer	Fall	Winter	Year
From	23.3	22.0	23.6	24.0	22.6
	(8.5)	(9.9)	(8.4)	(10.5)	(9.4)
To	23.3	22.0	23.6	24.0	22.6
	(7.8)	(7.8)	(7.9)	(5.1)	(7.1)
Polis weighted market access					
	Spring	Summer	Fall	Winter	Year
From	4.5	4.4	4.8	4.7	4.5
	(3.5)	(4.0)	(3.3)	(4.0)	(3.8)
To	4.7	4.6	4.4	5.1	4.5
	(2.9)	(3.1)	(3.1)	(2.6)	(2.8)

Table 1: Summary statistics of the main variables. The table shows arithmetic means with standard deviations in brackets. The number of observations is always the number of segments, which is 2,737. *To* and *From* statistics for mean market access are identical by construction for the corresponding variable.

results vary because the same values are assigned differently to individual segment points. To illustrate this point, Figure 2 shows the 10 percent closest segment points for four poleis in different parts of the Mediterranean. Given strong wind patterns, the correlation between *To* and *From* can even be negative. These variables are defined more technically below.

The polis weighted market access variables also count inverse distances, but weight segment points that can be reached by the *polis* variable of the destination. The table again shows means with standard deviations in brackets.

4 Empirical Results

Our empirical results address how the potential for interpolis trade helps explain the location where poleis have been attested to have emerged in that part of the world. The variety of climatic and soil characteristics around the Mediterranean and the Black Sea along with the forces of trade should allow, in principle, for a variety of settlement patterns. What did determine the location of urban settlements that we observe? The Mediterranean geography is conducive to urban development in the broadest of senses, economic, intellectual and political, which underlies the important intuition of Hicks (1969) and most recently of Matsuyama (2017).

We have already summarized, in section 2.5, the main predictions of our model. Succinctly, they are as follows. Under the assumption that there is trade among poleis, the modern spatial economics framework when populations of poleis are exogenous, that is labor supply is inelastic at every site, implies a system of simultaneous equations that wage rates or polis incomes obey; see section 2.2. An alternative assumption would be to assume free labor mobility, and thus infinitely elastic labor supply at every site, given a total population for the economy. In that case, spatial equilibrium rests on the notion that utilities are equalized across all sites; see section 2.3. Wage rates and populations are analytically well defined in terms of the left and right eigenvectors associated with the Perron-Frobenius eigenvalue of the non-symmetric matrix, $\tilde{\mathbf{A}}$, defined in (17), and whose elements, namely shipping costs, site amenities and productivities are exogenous.

For the exogenous populations case, as evident from (15), or its equivalent (12), this system of equations determines wage rates, given in addition shipping costs and geographic characteristics of the entire system. While structural estimation using those functional forms is possible, we opt for a reduced-form approach which is commonly used in this literature. This so-called market access approach, which was originally proposed by Harris (1954), is algebraically similar to (15) for the special case of $\sigma = 2$.[27]

Section 4.2 reports our empirical approach that uses both the poleis and the segments data sets, and aims at investigating the determinants of urban settlements in the classical Hellenic era. Section 4.3 reports a second group of regressions that aims at exploring the interdependence of the sizes of poleis, given their locations. Both groups of estimations are motivated by the same theoretical framework.

Roughly speaking, trade among urban settlements causes the likelihood of settlement at site s to be a function of productivities at all other sites and of shipping costs *to* and *from* all other sites. Shipping costs, that is the number of units of the locally produced good that must be shipped from site i to site j in order for one unit to arrive at site j, in units of the good shipped from i to j, are proxied by maritime travel times, which are directional and whose construction is discussed in sections 3.1 and 3.2 above and in further detail in the Appendix, section 7.1. The longer the travel times the higher the shipping costs.

There are several ways one could think of the network of voyage times. First, time distances proxy for the cost of shipping. They can also proxy for the ease of communication and therefore the transmission of information about political and military events. Because of its directed nature, ease of communication is also subject to seasonal weather patterns in prevailing winds.

While the popular market access approach informs patterns in urban settlements and their determinants, we have also experimented with additional empirical approaches. The spectral properties of the network's adjacency matrix are amenable to the usual interpretations as proxies of the network's centrality. We report briefly such results in section 4.2.3.

4.1 Spatial Expanse of the Ancient Hellenic System of Cities

Figure 3 depicts the sites of Neolithic settlements (that we know of) in the Hellenic world, that is the Mediterranean and the Black Sea regions and the sites of Hellenic poleis. A particular notable finding is that the average distance of Neolithic sites from the nearest sea coast of 31.1 km goes down to 9.973 km for the ancient poleis. So, the Hellenic urban system moved urban activity nearer to the sea; or, put alternatively, urbanization in the Hellenic world is associated with a leap towards the sea. Both pacification and the acquisition of maritime trade technology and skills must have been factors.

In more detail, working with Neolithic sites data we identified Neolithic sites in the vicinity of poleis in our data set and consider those that are: less than 1 km away, between 1 - 2km away, 2 - 3 km, 3-4 km, 4 -5 km, 5-10 km, 10-20 km, and 20-50 km. In our data set and among poleis with nonmissing size information, about one-third, that is 218 poleis of 636, have Neolithic sites within a 50 km radius from them. Comparison of size, for those poleis that have a Neolithic site within 50 km with those poleis that do not shows that the frequency distributions are quite similar. Regressions roughly along the lines of Size against dummy variables for existence of Neolithic sites over distance intervals 0-1km, 1-2km, 2-3km, 3-4km, 4-5km, 5-10km, 10-20km, and 20-50km yield highly statistically significant and positive results for the dummy variables for the intervals 2-3km, 3-4km, and 4-5km, but the adj-R^2 at 0.004 is very small. Therefore, there is some evidence of path dependence from the Neolithic era to the classical Hellenic era. This finding is confirmed when the counts of the number of Neolithic sites in the respective intervals instead of dummies are used, but the adj-R^2 barely increases to 0.005. The distance to the coast for Neolithic sites within 5 km of our poleis has a mean value of 31.1 km, with a minimum value of 1.3km and a maximum of 49.8km. It is much less concentrated than the distance to the harbor of the respective polis, which has a mean of 9.9 km.

4.2 Market Access Regressions

We present next regressions performed using both the poleis and the segments data sets, with the units of observation being segment points, $s \in \mathcal{S}$. These are 2,737 segment points that are located in regular distances along the coast of the Mediterranean and the Black Seas. See Appendix, section 7.1 for details. There is an important difference between the model and the empirical implementation: In the model, we do not account for unpopulated locations explicitly, even though they could be included simply by adding locations that are unsuitable for production or consumption (setting $A_i = 0$ or $u_i = 0$). In the data, we use a regular grid, which allows us to estimate the extensive and intensive margin of the location of poleis simultaneously. Variable $polis_s$ is defined as the count of the number of poleis that are within a distance of 10km of segment point s. Distances for this purpose are measured as Euclidian distances. Our empirical analysis using the concept of market access is motivated by (33). The theory predicts that the L_i's, the populations of sites at spatial equilibrium, depend on both consumer and producer market access, with the former having a relatively stronger effect than the latter, under the assumption of $\sigma > 1$. We interpret the definitions of the market access variables as reduced-form relationships.

4.2.1 Definitions

We compute various measures of market access based on inverse distance weighting, varying by season and direction of trade. For example, variable $MA\_spring\_to_s^{all}$ and variable $MA\_spring\_from_s^{all}$, where $d_{j,s}^{spring}$ is the travel time measured in spring showing market access *to* and *from* segment point s, defined as:

$$MA\_spring\_to_s^{all} = \sum_{j \in \mathcal{S}} [1/d_{j,s}^{spring}]; \;\; MA\_spring\_from_s^{all} = \sum_{j \in \mathcal{S}} [1/d_{s,j}^{spring}], \; s \in \mathcal{S}. \tag{34}$$

We use the other three seasons in addition to spring, as well as measures for the year average. In the notation below we occasionally omit the season classifier from these variables. These market access measures consider market access to all other segment points, regardless of the properties of these points. We compute alternative market access measures in which we weight target points. For example,

$$MA\_spring\_to_s^{polis} = \sum_{j \in \mathcal{S}} [polis_j/d_{j,s}^{spring}], \; s \in \mathcal{S}$$

computes market access to other segment points that also contain a polis.

$$MA\_spring\_to_s^{crop} = \sum_{j} [crop_j/d_{j,s}^{spring}], s \in \mathcal{S},$$

computes market access to segments weighted by crop growing suitability, where $crop_j$ is a variable measuring agricultural productivity at segment j.

Figure 2 illustrates these definitions, including the significance of the directionality of the imputed trading costs. The maps show the 10 percent closest segment points *To* and *From* for four notable poleis: Chalkis, Korinthos, Rhodos, and Syrakousai. The geographical reach of the market access of those large poleis is consistent with the historical evidence of "staggering series of luxury, strategic and basic commodities imported" by notable individuals [Kallet and Kroll (2020), p. 28-29].

Our segments-based regressions are of the form:

$$polis_s = \alpha + \beta MA\_spring\_to_s^{all} + \gamma \mathbf{X}_s + \epsilon_s \; s \in \mathcal{S}, \tag{35}$$

where matrix $\mathbf{X}_s$ contains a set of exogenous control variables, which are precipitation, temperature, elevation, standard deviation of elevation, malaria intensity and ruggedness; see Appendix, section 7.2 for their definitions. To account for spatial correlation we cluster standard errors at the level of a regular grid at 1×1 degree latitude and longitude. Variable MA_s^{polis}, defined in Equation (34), is potentially endogenous in OLS estimations of equation (35) because it depends on the endogenous variable $polis_j$ and hence there could be reverse causality (among other concerns). Variable MA_s^{all} is derived only from geography and not computed from any endogenous variables, and thus does not come with concerns over reverse causality. Hence we can use it as an instrument in a 2SLS estimation, where the first stage is:

$$MA_s^{polis} = \alpha_1 + \beta_1 MA_s^{all} + \gamma_1 \mathbf{X}_s + \epsilon_{1,s} \; s \in \mathcal{S}, \tag{36}$$

and the second stage is

$$polis_s = \alpha_2 + \beta_2 \widehat{MA_s^{polis}} + \gamma_2 \mathbf{X}_s + \epsilon_{2,s}. \tag{37}$$

This approach follows Bakker *et al.* (2021). We interpret $\hat{\beta}_2$ as the causal estimate of market access to poleis on the development of a segment point.

4.2.2 Empirical Results

Table 2, Panel A, reports estimations of equation (35) using the segment points data set, merged with the information from the poleis data set. As in all the following tables, the outcome variable is a count of poleis in the vicinity of a segment point, so coefficients can be interpreted as probability that a segment point develops an additional polis. All coefficients in this table are positive, and most are strongly statistically significant. We note that coefficients estimated using travel times for direction *To* other segment points are almost twice as large for the whole year than travel times *From* other segment points in this table. The coefficients for *To* are also large in every season except spring. These differences are particularly pronounced in fall and winter. The table also shows that there is some seasonal variation. We note a particularly strong coefficient for winter and travel direction *To* other segment points. In other words, market access matters most by providing imports in winter.

Table 2, Panel B, repeats this exercise, but computes market access *To* and *From* other poleis segment points weighted by the number of poleis within 10km. Hence, this panel considers market access to other cities rather than market access to coastal segments. Again, all coefficients in this table are positive, and all coefficients are strongly statistically significant. So, in this version of market access, only connections *To* other poleis count, and segment points without a polis carry a weight of zero in the market access computation. We notice that the coefficients *To* are larger than the coefficients *From* in all seasons and for the year as a whole. Again, having imports in winter is the largest of all coefficients.

We can get a sense of statistical significance of the difference between these coefficients by running regressions that include regressors that allow us to estimate two of these coefficients simultaneously. Using this method, we find for instance that the difference between *From* and *To* for the whole year in Panel B is statistically significantly at the 0.1 percent level. There is moderate seasonal variation, with again the strongest effect found for winter and travel direction *To* other poleis.

We repeat the exercise in Panel C, this time weighting segment points by their suitability for growing crops. Again, coefficients are positive and statistically significant, and larger for *To* for the year as a whole.

The OLS coefficients in Panel B suffer from a reverse causality endogeneity problem given that the location of a polis influences the location of other poleis that make up its market access. This is less a concern for Panel A and Panel C in which market access is purely a function of geographic variables that are exogenous to human construction. To address this concern, we adopt an instrumental variable procedure, using the exogenous market access to all segment points as an instrument for the endogenous market access to other poleis. The respective first stage of this exercise is shown in Panel A of Table 3, which shows that the

	(1) Spring	(2) Summer	(3) Fall	(4) Winter	(5) Year
Panel A: Market access to all segment points					
From MA_{all}	0.0249	0.0280	0.0059	0.0176	0.0176
	(0.0044)	(0.0044)	(0.0053)	(0.0037)	(0.0048)
To MA_{all}	0.0137	0.0312	0.0331	0.0507	0.0301
	(0.0053)	(0.0053)	(0.0047)	(0.0085)	(0.0057)
Panel B: Market access to poleis					
From MA_{polis}	0.112	0.103	0.112	0.082	0.095
	(0.0103)	(0.0096)	(0.0113)	(0.0092)	(0.0099)
To MA_{polis}	0.127	0.137	0.128	0.168	0.154
	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)
Panel C: Market access to crops					
From MA_{crop}	0.194	0.264	0.0378	0.179	0.146
	(0.042)	(0.046)	(0.043)	(0.040)	(0.045)
To MA_{crop}	0.088	0.213	0.248	0.230	0.187
	(0.046)	(0.053)	(0.046)	(0.052)	(0.048)
Observations	2,737	2,737	2,737	2,737	2,737

Table 2: OLS results, each coefficient comes from a separate regression. The outcome variable in all three panels is the sum of poleis within 10km of a segment point. Panel A shows market access to or from all segment points. Panel B shows market access to or from poleis. Panel C shows market access to or from crops. Control variables used in all regressions are: Precipitation, temperature, elevation, elevation (std), malaria, ruggedness. We cluster standard errors using a grid of 1×1 degree latitude and longitude.

	(1) Spring	(2) Summer	(3) Fall	(4) Winter	(5) Year
Panel A: First stage regressions					
From MA_{all}	0.338	0.390	0.275	0.336	0.359
	(0.042)	(0.0134)	(0.0263)	(0.0154)	(0.0145)
F-stat	282	138	845	179	110
To MA_{all}	0.272	0.318	0.334	0.389	0.299
	(0.0232)	(0.0238)	(0.0208)	(0.0305)	(0.0242)
F-stat	259	475	162	610	153
Panel B: Second stage regressions					
From $\widehat{MA}_{polis}$	0.0736	0.0719	0.0214	0.0523	0.0489
	(0.0106)	(0.0100)	(0.0180)	(0.0097)	(0.0121)
To $\widehat{MA}_{polis}$	0.0504	0.0980	0.0990	0.131	0.101
	(0.0162)	(0.0118)	(0.0114)	(0.0155)	(0.0136)
Observations	2,737	2,737	2,737	2,737	2,737

Table 3: 2SLS. Panel A shows coefficients from first stage regressions and Panel B shows coefficients from second stage regressions. The outcome variable Panel B is the sum of poleis within 10km of a segment point. Control variables used in all regressions are: Precipitation, temperature, elevation, elevation (std), malaria, ruggedness. We cluster standard errors using a grid of 1×1 degree latitude and longitude.

correlation between the instrument and the endogenous variable is positive and strong.

The second stage in Panel B shows the instrumented, causal effect of market access to other poleis on the probability that a location hosts a polis. Instrumented coefficients are smaller than in the corresponding OLS Table 2, which we attribute to the reverse causality mechanism being removed by the instrument. In terms of magnitudes, if we take the coefficient *From* for the whole year to be 0.05, it implies that increasing the market access by 1 (which would be the case if one polis is added at a distance of one) increases the probability that a site has an additional polis by 0.05. Also in this instrumented version, coefficients *To* are larger than coefficients *From*, about twice as large for the whole year. As in the OLS results, the largest coefficient is *To* in winter, suggesting that the most important aspect of market access is to provide imports in winter.

To assess the statistical significance of the difference between *From* and *To*, we run a version in which we include both instruments and both coefficients in the same set of regressions. We find the difference in the estimates to be statistically significant with a p-value of 0.0016. Using the same method, we find that the difference between Spring and Winter in the *To* regression has a p-value of 0.013. As further evidence for the importance of seasonal variation, we compute a variable that is the standard deviation of seasonal variation, as defined in equation (27). This is computed here as:

$$MA\_to\_std_s = \sqrt{\sum_{s \in season} (MA\_season\_to_s)^2},$$

where the set *season* includes Spring, Summer, Fall and Winter and variables are as defined above. When we include this variable to the estimations, such as those reported in Table 2, Panel B, Column (5) the estimated coefficient is 0.020 (0.011). The equivalent coefficient for a version for the *From* specification is 0.021 (0.010). Therefore, recalling the discussion in sections 2.4.1 and 2.4.2 we note that the empirical analysis supports the notion that seasonal fluctuations are more likely to lead to urban development.

In Table 4 we repeat the exercise from Table 3, Panel B, but vary market access by the size of poleis that can be reached. The table shows regression coefficients from 10 different market access coefficients. This table tests whether access to larger poleis is more valuable than access to smaller ones. Indeed, in columns (1) and (2) coefficients tend to increase in the size of polis that a segment point connects with, both in direction *To* and direction *From*. However, comparison of their magnitudes is influenced mechanically by the differences in the means and standard deviations of each of these different variables that measure market access, given the unequal number of poleis of each size. We run a version of Table 4 in which we standardise all ten poleis market access variables to mean zero and standard deviation one in columns (3) and (4). Again, the standardised coefficients increase with the size of the market access measure, suggesting that market access to and from larger poleis is more important than to and from smaller ones. Particularly large is the coefficient for *To* poleis of

	(1)	(2)	(3)	(4)
			Standardised	Standardised
	From	To	From	To
$\widehat{MA}_{polis\ size1}$	0.206	0.473	0.197	0.332
	(0.051)	(0.0581)	(0.0482)	(0.0407)
$\widehat{MA}_{polis\ size2}$	0.166	0.293	0.192	0.276
	(0.041)	(0.0455)	(0.0472)	(0.0427)
$\widehat{MA}_{polis\ size3}$	0.403	0.683	0.203	0.309
	(0.101)	(0.104)	(0.0511)	(0.0471)
$\widehat{MA}_{polis\ size4}$	0.463	0.871	0.207	0.297
	(0.121)	(0.126)	(0.0542)	(0.0430)
$\widehat{MA}_{polis\ size5}$	0.581	1.939	0.212	0.561
	(0.142)	(0.370)	(0.0519)	(0.107)
Observations	2,737	2,737	2,737	2,737

Table 4: 2SLS: Second stages for poleis of different size. The outcome variable is the sum of poleis within 10km of a segment point. Control variables used in all regressions are: Precipitation, temperature, elevation, elevation (std), malaria, ruggedness. We cluster standard errors using a grid of 1×1 degree latitude and longitude.

size 5. We note again that for poleis of all sizes, connecting *To* them is more valuable than being connected *From* them, both in the standardised and non-standardised version.

Next we examine the significance of observable physical attributes of poleis, such as evidence of fortifications in the form of walls. Existence of walls is controlled for in linear regressions by computing market access to poleis with walls and, separately, to poleis with no walls. We first include market access to poleis with and without walls in a *To* regression. Coefficients are 0.164 (0.032) for market access To poleis with walls and 0.142 (0.054) for market access To cities without walls. Both are strongly statistically significant, and access to cities with walls matters slightly more (but not statistically significantly differently so). Corresponding values for the equivalent regression From poleis with and without walls are 0.1581 (0.0388) from Poleis with walls and 0.003 (0.0647) from Poleis without. Here only market access from cities with walls is statistically significant, and the difference between both coefficients is highly statistically significant. These regressions confirm that market access to larger cities is more important; interestingly this difference is much more pronounced for the *From* specifications.

We now examine the significance of such political attributes as polis membership in the

Delian League or in federations, known as *Koinon.* Instrumental variable regressions with two different market access variables, one To poleis that are members of the Delian League and those that are not, show that the former has a positive and the latter a negative effect on a site. In the *To* regressions, coefficients are 0.261 (0.070) for market access To Delian League members and -0.109 (0.089) To non-members. *From* coefficients are 0.515 (0.177) and -0.564 (0.238) respectively. This confirms the notion that in the long run, membership in the Delian League, the institutional structure of Athenian hegemony, complete with adoption of standards and payments of taxes in kind and via seignorage, had a greater impact on urban development, possibly because it helped divert resources away from non-Delian League members and in favor of members. Also, it could reflect the impact of the military side of Delian membership on reducing trading costs via facilitating maritime shipping.

Another political attribute is a polis' membership to a *Koinon*, which means "common" and designates membership to a regional Federation prior to 330 BCE, of which several existed [see MacKil (2013).] Interdependence among them was not ruled by a hegemon, unlike Athens and the Delian League; outcomes were quite decentralized in comparison. Instrumental variable regressions with two different market access variables, one to poleis that are members of a Koinon and those that are not show that the coefficients in the *From* regression are not statistically significant (8.352 (16.79)) and (-1.434 (1.968)) respectively. In the case of the *To* specifications, the former is negative and not statistically significant at five percent (-0.433 (0.240)), while the latter is positive and statistically significant (0.201 (0.043)). This agrees with the notion that belonging to *Koinon* membership is a rather "loose" concept, compared to the Delian League.

4.2.3 Robustness Check

One of the main points we want to emphasize with the above empirical analysis is the role of asymmetric trading costs, which we explore by means of the canonical market access approach. As a robustness check, we consider in this section measures of centrality, which are based on the eigenvectors defined by (20) and (21). They, too, aggregate the impact of the entire urban system on urbanization at each segment. It is thus reasonable to compare

	(1) Spring	(2) Summer	(3) Fall	(4) Winter	(5) Year
Authority (To) Centrality	0.119	0.167	0.128	0.161	0.154
	(0.0244)	(0.0227)	(0.0218)	(0.0242)	(0.0235)
Hub (From) Centrality	0.127	0.162	0.110	0.134	0.130
	(0.0240)	(0.0221)	(0.0219)	(0.0230)	(0.0231)

Table 5: OLS results, each coefficient comes from a separate regression. The outcome variable is the sum of poleis within 10km of a segment point. Control variables used in all regressions are: Precipitation, temperature, elevation, elevation (std), malaria, ruggedness. We cluster standard errors using a grid of 1×1 degree latitude and longitude.

their performance with those of market access. 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. 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.

We compute the authority and hub centrality for a matrix whose entries are given by the inverse of the distance between each segment pair: $\left[\tau_{ij}^{-1}\right]$. The results with standardized centralities as the explanatory variables are reported in Table 5, which corresponds to Panel A of Table 2. All of the centralities are highly significant, and the authority centrality (which is linked to the left eigenvector of the above matrix), always has a larger coefficient than the hub centrality (which is linked to the right eigenvector) for all seasons other than spring. Although those two centralities are highly correlated, except for in spring, when they are included in regressions together, authority centrality has positive and significant coefficients, but hub centrality is not always so. Moreover, in such specifications, the coefficient of authority centrality is always larger than the respective one for hub centrality, and equality of coefficients for those two variables is rejected. In view of their definitions, *To* is akin to authority centrality, and *From* to hub centrality. Therefore, our results corroborate our claim based on market access measures that for urbanization in each segment, *To*, being better connected from other segments, is more important than *From*, being better connected to other segments.

4.3 Poleis-based Regressions

Our empirical analysis so far focuses on the impact of trade across the ancient Greek cities, which in turn sustains urban development. Local productivity characteristics near sites occupied by poleis are critical in sustaining production of diverse goods that are traded within the system of cities. Naturally, we would expect that such local characteristics would provide for urban economies under autarky. The model of section 2 is predicated upon interpolis trade. Absent trade, the local real wage rate is determined but not the size of settlement.

In the remainder of this section we discuss results with regressions of polis size against a variety of local characteristics which proxy for A_i. We do so with a word of caution. Emergence of poleis at particular sites are not random events. The regressions of polis sizes against own characteristics do not control for that. Our market access regressions, reported above, do so in connection of presence of poleis in the vicinity of each segment. In future work, we will study more closely the consequences of ignoring this selection bias.

Following Henderson *et al.* (2018) we consider the crop suitability variables, ruggedness variables, malaria index, trade-related variables such as distance to coast, distance to nearest harbor, distance to permanent river, distance to temporary river, and distance to nearest lake. Regarding crop suitability, we tried barley suitability, summer wheat suitability, and winter wheat suitability, which we think are most pertinent for that part of the world. All these variables are measured in terms of their percentiles in the sample, although results (not shown) are very similar if we use the actual values or their logs.

4.3.1 The Role of Own Local Characteristics

Referring to Table 6, columns 1 and 2 report regressions with size and columns 3 and 4 with log of size against the entire complement of own local characteristics, with as well as without historical region fixed effects. The ruggedness and malaria variables are not statistically significant in these regressions, while ruggedness within 20km has a negative sign in univariate regressions without fixed effects. Greater ruggedness is associated with smaller polis size but explains very little of the variance. Of the crop suitability variables, barley and summer wheat are statistically significant, with the former having a positive sign and the latter a negative one. However, they explain very little of the variance in univariate regressions.

Coming next to the trade-related variables, distance to the coast and distance to the nearest harbor, are both always significant with a positive and a negative coefficient, respectively, even with fixed effects. Distance to a permanent river and a temporary river both have positive and significant coefficients in univariate regressions without fixed effects but only distance to temporary river retains its significance in the multivariate regressions, reported in Table 6, though only in the case of no fixed effects. Finally, distance from a harbor is a disadvantage, in that its respective coefficient is negative and significant, but distance from the coast is an advantage. We may conclude that ease of navigation is an advantage but exposure to the coast a disadvantage perhaps because of security considerations.

In sum, when all explanatory variables are included in the multivariate regressions, the geography and crop suitability variables perform best. Including historical region fixed effects improve the adj-R^2 from 0.083 and 0.060 to 0.275 to 0.250, for size and log of size, respectively. Clearly, the regional dimension is important. Finally, comparing the R^2 of these regressions with the corresponding market access ones suggests suggests that interpolis trade, as proxied by such explanatory variables as $\widehat{MA}_s^{polis}$, is a significant determinant of urbanization in the ancient Hellenic world.

5 Conclusions

The main focus of this research is the importance of trade for polis sizes and patterns of urbanization in the ancient Hellenic world. This is a setting in which we can measure asymmetries in transport costs due to prevailing strong winds. Our findings establish such importance using market access regressions. These regressions are motivated by a standard model of international trade. It is particularly noteworthy that asymmetries in shipping costs are quantitatively important, and robustly so across a range of different regressions.

	Size		ln(Size)	
	(1)	(2)	(3)	(4)
Malaria	-0.000581	-0.00297	-0.000567	-0.00169
	(0.00209)	(0.00202)	(0.000903)	(0.000886)
Ruggedness	0.00369	0.00499	0.000776	0.00197
	(0.00330)	(0.00409)	(0.00139)	(0.00177)
Distance to coast	0.0121	0.00676	0.00541	0.00324
	(0.00332)	(0.00329)	(0.00146)	(0.00143)
Distance to harbor	-0.0120	-0.00881	-0.00513	-0.00423
	(0.00324)	(0.00298)	(0.00145)	(0.00133)
Distance to permanent river	0.00249	-0.000898	0.0000727	-0.000560
	(0.00282)	(0.00380)	(0.00121)	(0.00169)
Distance to temporary river	0.0108	-0.00187	0.00388	-0.000543
	(0.00267)	(0.00362)	(0.00116)	(0.00159)
Suitability: Barley	0.00977	0.0104	0.00343	0.00449
	(0.00470)	(0.00577)	(0.00197)	(0.00245)
Suitability: Summer wheat	-0.00955	-0.00773	-0.00395	-0.00320
	(0.00454)	(0.00607)	(0.00188)	(0.00251)
Suitability: Winter wheat	-0.00305	-0.00549	-0.00109	-0.00273
	(0.00277)	(0.00376)	(0.00115)	(0.00158)
Historical Region FE		X		X
Observations	620	620	620	620
Adjusted R2	0.083	0.275	0.060	0.250

Table 6: OLS regression for poleis size against own local characteristics. The outcome variable is the size of the poleis (discrete values from 1 to 5). All explanatory variables are scaled to be the percentile (ranging from 0 to 100) of the actual values. Results are similar if we use the actual or log of actual values of the explanatory variables. The constant is included not shown. Robust standard error in parentheses.

Import market access, *To* a polis, is consistently a stronger predictor of the location of a polis than export market access, *From* a polis. The quantitatively most important market access variable we find is importing in winter. We also document that seasonal variation in market access also matters for the emergence of early cities. These facts about trade may be true more generally.

In addition, our research provides new insights into the factors that determine the location of cities in ancient Greece. We provide facts on the role played by crop suitability variables and trade-related variables associated with a polis' own vicinity as well as those of the other poleis. Our key trade-related measures go much beyond the commonly used local geography-based trade-related covariates, thus providing new insights on the geographic fundamental factors that shaped the urban geography of ancient Greece.

6 References

Abulafia, David. *The Great Sea: A Human History of the Mediterranean.* 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.

Akpinar, Adem, Bilal Bingölbali, and Gerbrant Ph Van Vledder. 2016. "Wind and Wave Characteristics in the Black Sea Based on the SWAN Wave Model Forced with the CFSR Winds." *Ocean Engineering.* 126:276-298.

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.

Ashraf, Quamrul, and Stelios Michalopoulos. 2015. Climatic Fluctuations and the Diffusion of Agriculture." *Review of Economics and Statistics.* 97(3):589-609.

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.

Behrens, Kristian and Pierre Picard. 2011. "Transportation, Freight Rates, and Economic Geography." *Journal of International Economics*. 85:280-291.

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.

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

Bosker, Maarten, Eltjo Buringh, and Jan Luiten van Zanden. 2013. "From Baghdad to London: Unraveling Urban Development in Europe, the Middle East, and North Africa, 800-1800". *Review of Economics and Statistics*. 95(4):1418-1437.

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

Brandt, H., and K. Hochkirch. 1995. "The Sailing Properties of the Hanse Cog in Comparison with Other Cargo Sail Ships." *Ship Technology Research.* 42(1):3-21.

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

Casson, Lionel. *Ships and Seamanship in the Ancient World*. 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. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4081451.

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.

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

Finley, M. I. 1999. *The Ancient Economy.* Berkeley, CA: University of California Press.

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.

Harris, Edward M., David M. Lewis, and Mark Woolmer. 2016. *The Ancient Greek Economy: Markets, Households and City-States.* Cambridge: Cambridge University Press.

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

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

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

Horn, Roger A., and Charles R. Johnson. 2013. *Matrix Analysis.* Second Edition. Cambridge: Cambridge University Press.

Ioannides, Yannis M. “Location Decisions under Uncertainty and the Urban Model.” *Economics Letters.* 11:291-295.

Jackson, Matthew O., and Stephen Nei. 2015. “Networks of Military Alliances, Wars, and International Trade.” *PNAS.* 112(50):15277–15284.

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

Kulatilaka, Nalin, and Stephen Gary Marks. 1988. “The Strategic Value of Flexibility: Reducing the Ability to Compromise.” *American Economic Review.* 78(3):574–580.

MacKil, Emily. 2013. “The Greek *Koinon.*” In: Fibiger Bang, Peter, and Walter Scheidel. Eds. *The Oxford Handbook of the State in the Ancient Near East and Mediterranean.* Oxford Handbooks Online. Oxford.

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

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

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

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

Plato. Phaedo. 109a-109b.

http://www.perseus.tufts.edu/hopper/text?doc=

Perseus%3Atext%3A1999.01.0170%3Atext%3DPhaedo%3Asection%3D109b

Poulain, Pierre-Marie. Milena Menna1, and Elena Mauri. 2012. Surface Geostrophic Circulation of the Mediterranean Sea Derived from Drifter and Satellite Altimeter Data." *Journal of Physical Oceanography.* 973–990.

Redding, Stephen J. 2020. "Trade and Geography." NBER Working Paper 27821. September.

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.

Soukissian, Takvor, Aristides Prospathopoulos, Maria Hatzinaki, and Maria Kabouridou. 2008. "Assessment of the Wind and Wave Climate of the Hellenic Seas Using 10-Year Hindcast Results." *The Open Ocean Engineering Journal.* 1:1-12.

https://benthamopen.com/ABSTRACT/TOOEJ-1-1

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.

Whiteright, Julian. 2011. "The Potential Performance of Ancient Mediterranean Sailing Rigs." *The International Journal of Nautical Archaeology.* 40:2–17.

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.

Notes

[1]The paper follows the convention in the classics literature and adopts Greek spelling, so that poleis is the plural for polis.

[2]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 (2005).

[3] 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.

[4]Specifically, Bonnier *et al.*, 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 did local producers choose to plant olives instead of sowing grains, when the demand for this basic foodstuff seemed to be increasing? Their explanation is that 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.

[5]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 but also as a luxury good associated with sports in

the gymnasia.

[6]See Kallet and Kroll (2020) for a discussion of 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.*

[7]See MacKil (2013) for a discussion of the role of *Koinon.*

[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 *ibid.*, Ch. 3 in connection with Aristotle's view of individuals as "political animals."

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

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

[11]Alternatively, poleis may be defined in terms of a publicly financed public good. See Chen and Ioannides (2022) for such a model with a fixed number of goods.

[12]The alternative assumption, namely that an endogenous range of intermediate goods are produced in every site under conditions of monopolistic competition and increasing returns to scale provides an equivalent conceptual framework for our purposes [Redding and Rossi-Hansberg (2017)].

[13]Allen, Arkolakis and Takahashi (2019) is the latest treatment of quasi-symmetric trading costs, that is, the case where $\tau_{ij} = \tau_{i}^{\cdot}\tilde{\tau_{ij}}\tau_{j}^{\cdot}$, where $\tilde{\tau}_{ij}$ are the elements of a symmetric matrix and $\tau_{i}^{\cdot}, \tau_{j}^{\cdot}$ are the elements of positive vectors.

[14]Some additional properties of this solution are as follows. By Theorem 8.2.8 of Horn and Johnson (2013), p. 526, $\bar{\Omega}$ is an algebraically simple eigenvalue of $\tilde{\mathbf{A}}$, the Perron-Frobenius maximal eigenvalue, and its corresponding right eigenvector $\tilde{\mathbf{x}}$ is unique and positive. If $\tilde{\mathbf{x}}$ is normalized by $\sum_i \tilde{x}_i = 1$, then the corresponding left eigenvector $\tilde{\mathbf{y}}$, which is also unique and positive, may be normalized by $\sum_i \tilde{x}_i\tilde{y}_i = 1$. Furthermore, by *ibid.,* Theorem 1.4.7 (b), p. 78, every left eigenvector of $\tilde{\mathbf{A}}$ other than $\tilde{\mathbf{y}}$ is orthogonal to all right eigenvectors of $\tilde{\mathbf{A}}$ other than $\tilde{\mathbf{x}}$, and likewise, every right eigenvector of $\tilde{\mathbf{A}}$ other than $\tilde{\mathbf{x}}$ is orthogonal to all left eigenvectors of $\tilde{\mathbf{A}}$ other than $\tilde{\mathbf{y}}$. This implies that $\lim_{m\to\infty}\left(\frac{\tilde{\mathbf{A}}}{\bar{\Omega}}\right)^m \to \tilde{\mathbf{x}}\tilde{\mathbf{y}}^T$, which is a positive rank-one matrix.

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

[16] In fact, sailing times across the poleis in the data are also available on a monthly basis.

[17] Thanks go to Theodore Papageorgiou for clarifying this matter. See also Brancaccio, Kalouptsidi and Papageorgiou, *op. cit.*

[18]See Kulatilaka and Marks (1988) who show in a deterministic model that under certain conditions the

value of flexibility can be negative.

[19]This is conceptually similar to the treatment of Redding and Rossi-Hansberg (2020), section 3.5.

[20]Ioannides (1983) offers some analytical arguments about the tradeoffs associated with fluctuating transportation costs, though they are obtained in a different urban setting and imply that city size increases with uncertainty.

[21]This is similar to the convexity property of indirect utility with respect to prices. However, this is not the end of the story because as Samuelson (1972) points out, the ultimate comparison is in terms of the equilibrium outcomes that prevail with random prices, given all exogenous parameters: "the consumer does benefit from feasible price stability."

[22]We thank Josiah Ober for emphasizing this point.

[23]C.f. Fujita *et al.* (1999).

[24]An alternative formulation of the model in terms of quasi-linear quadratic preferences does not allow us to obtain empirically relevant predictions of greater clarity.

[25]Lest it be thought that (33) is specific to our model, we note that it is featured in Equation (2), Redding and Sturm (2008), although their empirical application does not involve asymmetric trading costs.

[26]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.

[27]The point that the Harris market access formulation is conceptually and analytically close to the implications of modern economic geography was first made forcefully by Paul Krugman.

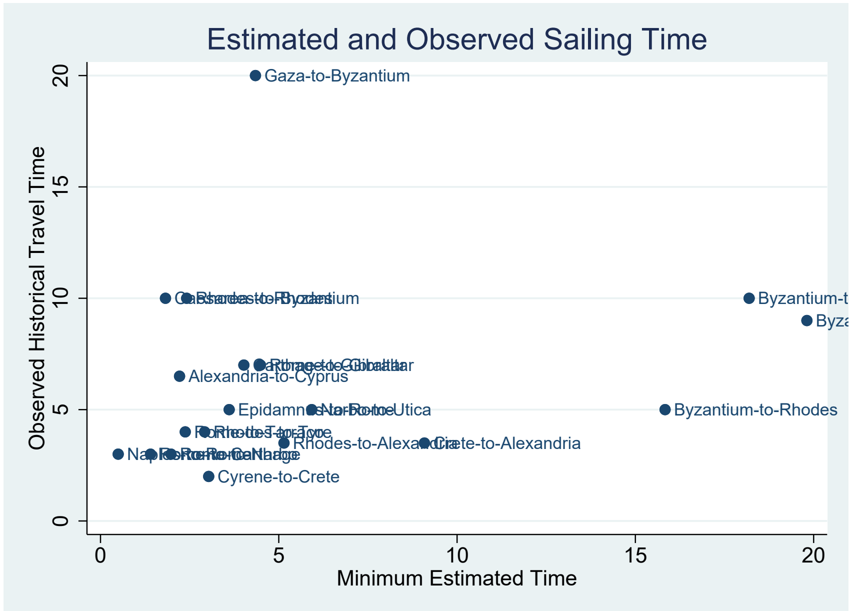


Figure 1: Travel Times: Historical Observed vs. Estimated

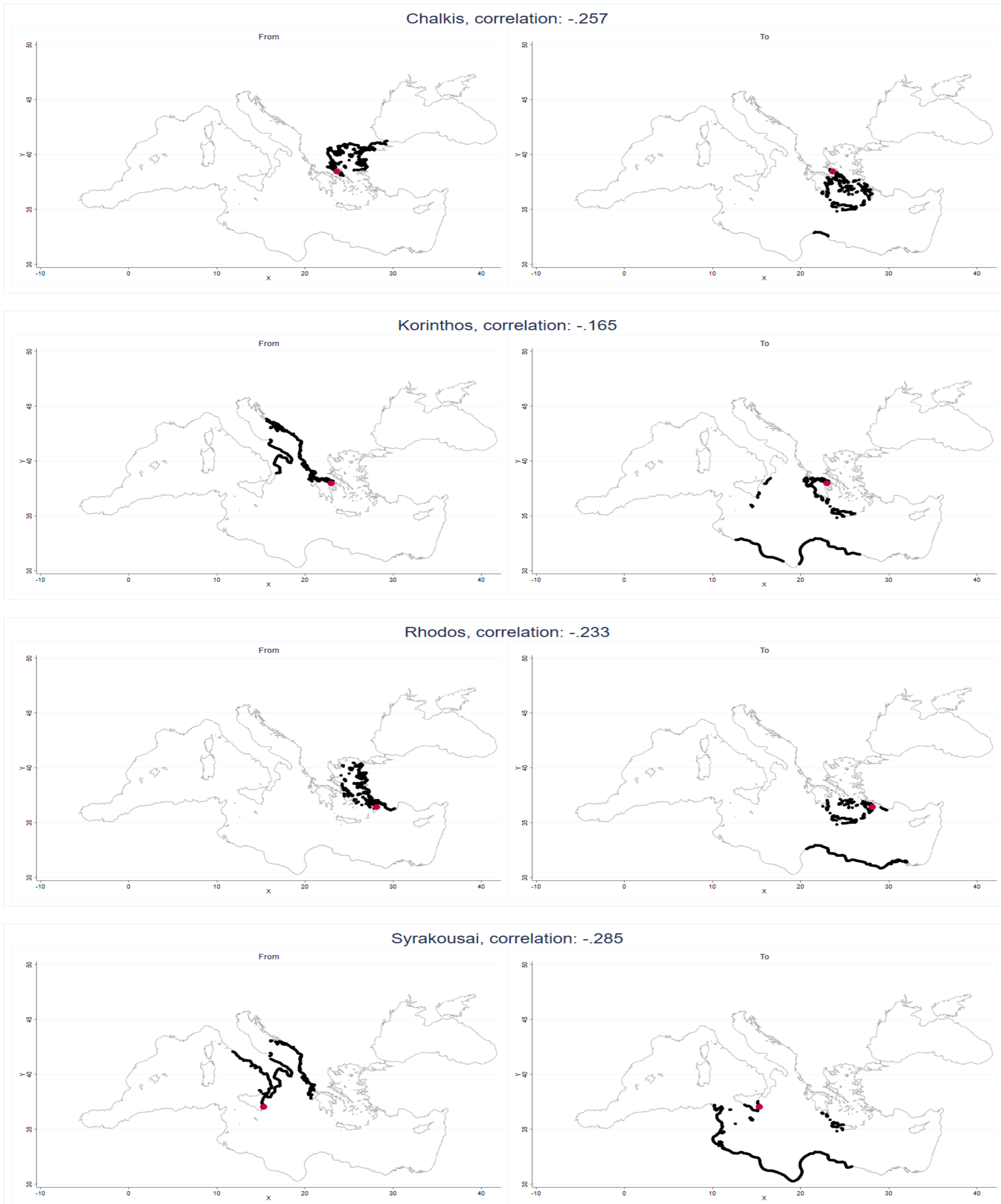


Figure 2: These maps show the 10 percent closest segment points *To* and *From* four example poleis. The plots also show the correlation between the *To* and *From* market access for segment points.

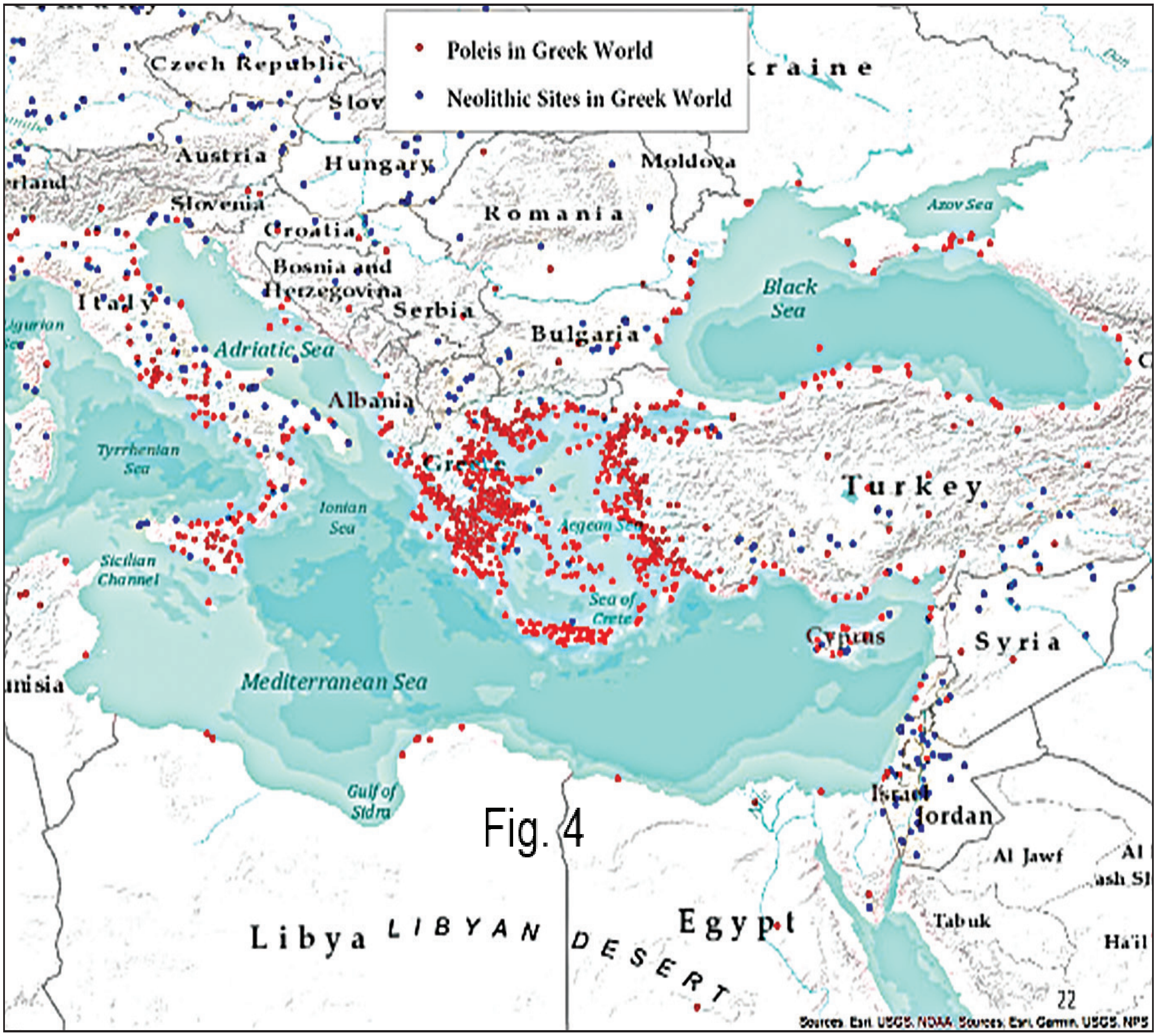


Figure 3: Neolithic Sites and Poleis in Hellenic World

7 Appendix: Data Construction

7.1 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 and 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.

The wind data come from NASA site

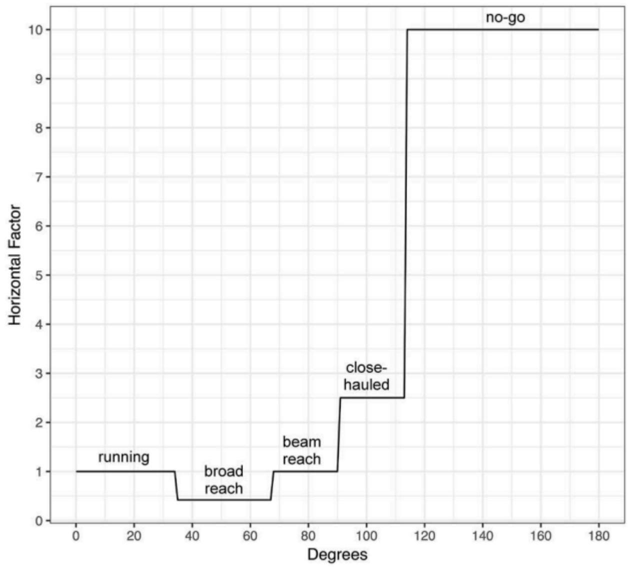
https://podaac.jpl.nasa.gov/dataset/CCMP_MEASURES_ATLAS_L4_OW_

L3_5A_MONTHLY_WIND_VECTORS_FLK

Any land has a "No Data" value, and thus we do not allow inland travel. The cost surface is not directed. To get the directed sailing time, a horizontal factors raster (wind direction in this case) is used, which applies a multiplier to the cost surface value, based on the angle between the direction of motion and the wind direction.[28] It allows the cost to be different when travelling in different directions. See Figure 4 for the exact value of the horizontal factor. 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.

The correlation between actual historical sailing times for some select routes listed with the calculated sailing times using the above network is demonstrated in Figure 1. Regressing minimum (among all seasons) estimated against historical observed minimum voyage times

Figure 4: Horizontal Factor Table



Source: Figure 4 of Alberti (2018)

has an adjusted R^2 of 0.425

7.2 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