Roman Transport Network Connectivity and Economic Integration *

Matthias Flückiger, Erik Hornung, Mario Larch, Markus Ludwig, and Allard Mees †

Final version

Abstract

We show that the creation of the first integrated multi-modal pan-European transport network during Roman times influences economic integration over two millennia. Drawing on spatially highly disaggregated data on excavated Roman ceramics, we document that contemporary interregional trade was influenced by connectivity within the network. Today, these connectivity differentials continue to influence integration as approximated by cross-regional firm investment behaviour. Continuity is partly explained by selective infrastructure routing and cultural integration due to bilateral convergence in preferences and values. We show that our results are Roman-connectivity specific and do not reflect pre-existing patterns of exchange using pre-Roman trade data.

Keywords: Economic integration, Roman trade, transport network connectivity, business links, cultural similarity

JEL classification codes: F14, F15, F21, N73, R12, R40, O18

*We thank Alina Bartscher, Sascha O. Becker, Michael Burda, Sebastian Braun, Alan de Bromhead, José De Sousa, Hartmut Egger, Richard Franke, David Jacks, Krisztina Kis-Katos, Udo Kreickemeier, Miren Lafourcade, Claire Lelarge, Stephan Maurer, Eric Melander, Luigi Pascali, Fabian Wahl, David Weil, Niko Wolf, and seminar participants at Bayreuth, HU Berlin, Bonn, TU Braunschweig, Brown, Cologne, George Mason, Goethe Frankfurt, Göttingen, Hamburg, Hong-Kong University, Konstanz, Mannheim, Paris-Sud, Stellenbosch, WU Vienna, the 2nd Workshop on Geodata and Economics in Hamburg, the Canadian Economics Association Annual Meeting 2019, and the 2020 RIDGE Virtual Forum for excellent feedback. We are grateful to Giovanni Federico for making the price correlation data available, Ömer Özak for kindly sharing data on the Human Mobility Index and Rome2rio.com for providing us with data on travel times. We also thank Kathrin Bank, Lukas Diebold, Tim Kleinlein, Leandro Navarro, Oliver Parker, and Antonia Weddeling for excellent research assistance. Hornung acknowledges that his research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. We declare that we have no relevant or material financial interests that relate to the research described in this paper.

†Flückiger: University of York, York, YO10 5DD, United Kingdom (matthias.flueckiger@york.ac.uk); Hornung: CAGE, CEPR, CESifo, ECONtribute, and University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany (hornung@wiso.uni-koeln.de); Larch: CEPII, CESifo, ifo, GEP, and University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany (mario.larch@uni-bayreuth.de); Ludwig: CESifo and Technical University of Braunschweig, Spielmannstraße 9, 38106 Braunschweig, Germany (markus.ludwig@tu-braunschweig.de); Mees: Romano-Germanic Central Museum - Leibniz Archaeological Research Institute, Ernst-Ludwig-Platz 2, 55116 Mainz, Germany (mees@rgzm.de)

1 Introduction

Large-scale transport infrastructure projects shape connectivity patterns and determine the distribution of economic activity across space by altering the costs of exchange. Changes in connectivity may have additional long-lasting consequences for connected regions because repeated interactions reduce information frictions and increase cultural integration. While many studies investigate the consequences of changing trade costs and transport infrastructure for levels of integration, we know surprisingly little about the potential causes of systematic differences in bilateral transport connectivity and information frictions between regions. For example, one of the largest infrastructure projects in history, the Belt and Road initiative, follows the Silk Road over long stretches. Along this ancient trade corridor goods, ideas, and cultural values have been exchanged over millennia. If new infrastructure projects follow existing patterns of economic integration, transport costs as well as informal barriers to integration—such as cultural differences—may be determined by historical economic integration. Hence, policy makers and economists need to be aware of the history of bilateral exchange and the concurrent integration of attitudes and tastes when evaluating infrastructure projects and regional policies and when discussing the optimal allocation of infrastructure resources.

This paper argues that the first pan-European multi-modal transport network—created by the Romans—had fundamental and lasting effects on the intensity of interregional (socio-) economic exchange. The unprecedented reach of the integrated network, combined with technological and institutional progress, dramatically reduced transport costs and changed the pattern of interregional trade in Western Europe. Regions better connected within the network started interacting more intensely and this pattern continued long after the collapse of the Roman Empire. The continued (socio-)economic interaction led to the convergence of preferences and values and thereby reduced information frictions. High similarity in these aspects, in turn, facilitates investment flows. Based on the arguments outlined above, we hypothesise that variation in connectivity within the Roman transport network determined historical trade flows and influences the intensity of cross-regional firm ownership today.

To empirically assess the validity of these hypotheses, we analyse the relationship between Roman connectivity and (socio-)economic integration at various points in time. We divide Western Europe into grid cells of 0.5×0.5 degrees and determine for each pair of cells how well it is connected within the multi-modal Roman transport network.[1] This network is a collection of numerous segments—sections of sea, river, or road—which differ in length and associated mode of transport. Based on Diocletian's Edict on Prices of 301 CE, a contemporary and widely used source, we determine Roman-technology-driven differences in freight rates across transport modes. Combining information on network and freight-rate differentials, we identify the least-cost path between any two grid cells that are connected to the network. The cost associated with shipping goods along this optimal path (referred to as *Roman effective distance*) constitutes our measure of connectivity within the Roman transport network. To isolate the Roman-era-specific aspects of this measure, we control for geodesic distance and further geographic controls

[1] In the remainder of this paper, we use the terms 'grid cell' and 'region' interchangeably.

throughout our empirical analysis.

Our first contribution is to document that variation in Roman effective distance influenced the intensity of trade during Roman times. To this end, we draw on geocoded information on more than 242,000 excavated potsherds of Roman fine tableware collated in the hitherto underexploited Samian Research database (Römisch-Germanisches Zentralmuseum in Mainz). A unique feature of the mass-produced and widely used ceramic tableware—subsequently referred to as 'terra sigillata'—is that production sites (i.e., the origins of the tableware) are precisely identifiable. Combined with information on the location of archaeological excavation sites (i.e., the destination of the terra sigillata), this allows us to aggregate the number of finds to the grid-cell-pair level and thereby capture interregional trade volumes within Western Europe during the Roman era. The possibility to trace terra sigillata from origin to destination, combined with the fact that it was traded throughout the entire Roman territory, makes them ideal goods to study long-distance trade in the first European-wide integrated market.

We empirically estimate the relationship between historical trade shares and Roman effective distance employing a Poisson Pseudo-Maximum-Likelihood (PPML) regression approach that accounts for heteroskedasticity and takes into account information contained in zero trade flows (see Santos Silva and Tenreyro, 2006; Eaton, Kortum and Sotelo, 2013; Barjamovic et al., 2019; Sotelo, 2019). To control for unobserved origin- and destination-specific effects, we include origin and destination fixed effects. The results document that Roman effective distance strongly influenced the volume of interregional trade. A one percent increase in Roman effective distance reduces trade by 2.9% when only controlling for origin and destination fixed effects. When we account for geodesic distance and other geographical disparities to isolate the Roman-era-specific part of effective distance, the point estimate remains statistically significant and economically sizeable, implying an elasticity of Roman effective distance of -1.5. This elasticity is close to estimates for other historical periods (see, e.g., Barjamovic et al., 2019; Donaldson, 2018; Wolf, 2009) and larger in absolute terms than estimates of modern-day distance elasticities (see, e.g., Disdier and Head, 2008).

As a second contribution, we provide evidence that Roman-era-specific transport network connectivity continued to influence the geography of trade at least until the advent of steam power and new transport technologies during the Industrial Revolution. Trade is proxied by the degree of interregional price correlation over the period 1208–1790 as well as differences in the timing of onset of the Black Death (1347–51).

Our third contribution is to document that differential connectivity within the Roman network influences the spatial pattern of firm ownership today, despite the fundamental changes in relative transport costs that occurred since the advent of railways and air travel. Drawing on geocoded firm-level data from the Bureau van Dijk's Orbis database, we show that greater connectivity during the Roman Empire intensifies cross-regional parent-subsidiary connections. In our preferred regression specification, which accounts for geodesic distance, geographical factors and home bias, ownership connections decline by 0.4% as Roman effective distance increases by 1%. This finding highlights that today's pattern of bilateral economic integration in Western Europe is (partly) determined by infrastructure routing decisions made 2,000 years ago. Motivated

by the observation that foreign direct investment is an important transmission channel of business cycles (Cravino and Levchenko, 2017), we extend our analysis and show that the effect of Roman connectivity on firm ownership is also reflected in business cycle integration. As proxy for integration we use correlation in night-time luminosity growth.

Our fourth contribution is to investigate potential mechanisms that link variation in connectivity within the ancient transport network to cross-regional firm investment behaviour today. Guided by recent studies (discussed below), we focus on two mechanisms: persistence in transport infrastructure connectivity and cultural convergence. We first show that regions better connected within the Roman transport network continue to be more closely linked within today's transport network, particularly the passenger network. This suggests that regions with stronger ancient connectivity were connected more directly when new transport technologies (e.g., railways, aeroplanes, and highways) became available. The persistence in infrastructure connectivity can explain a substantial part (36%) of the Roman-era-specific effect on cross-regional firm ownership. Second, we show that the effect of Roman-transport-network connectivity can partially be explained by network-induced similarity in preferences and values. Greater connectivity between regions increases similarity in preferences and values as reported in the Global Preferences Survey (GPS, Falk et al., 2018) and the European Values Study (EVS, EVS, 2016). This mechanism can account for 18% of the Roman-transport-network effect on firm ownership. Combined, the two mechanisms absorb around 50% of the ancient connectivity effect. The lack of detailed bilateral data prevents us from testing further specific mechanisms, such as greater genetic similarity resulting from migration. However, we use the Social Connectedness Index (SCI, Bailey et al., 2018*a*) as an aggregate index that captures the realisation of many conceivable mechanisms. This measure of the interregional intensity of social ties may be (loosely) interpreted as an index of revealed similarity. Once we account for the SCI, Roman transport network connectivity ceases to have explanatory value for today's ownership links.

The identifying assumption underlying our estimation strategy is that, conditional on control variables, Roman effective distance captures Roman-era-specific variation in transport network connectivity that is not correlated with unobserved factors that influence integration. A particularly pressing concern is that connectivity within the Roman transport network could be endogenous to pre-existing cultural and economic exchange. We alleviate this concern by empirically documenting that Roman effective distance does not explain pre-existing patterns of socio-economic integration. These findings are based on a newly compiled database of more than 7,000 prehistoric artefacts with precisely identifiable origins and destinations. We additionally show that connectivity within the not yet existing transport network has no effect on cultural diffusion, as measured by Neolithic or Bronze Age burial traditions.

Another related concern is that the path of Roman roads could have been influenced by economic or geographic factors. Historical evidence, however, indicates that the routing of the road network was primarily influenced by military strategic considerations and disregarded local (socio-)economic and geographic conditions (see, e.g., Temin, 2012; Laurence, 2002; Davies, 1998). Based on this historical evidence, we develop an instrumental variable approach in which we replace the actual road network with a hypothetical network that connects the capital Rome

to the locations of Roman battlefields using a Gabriel graph. This creates a network that consists only of straight-line road segments and allows for a rapid movement of troops. The results of the IV procedure confirm our main findings: Roman effective distance deters economic integration both in the past and today.

A further worry is that Roman effective distance partially captures the effects of geographic connectivity. To assuage this concern, we show that estimates remain stable when we account for a variety of geography-based least cost measures as well as a wide range of geographic aspects. Additionally, we conduct a falsification exercise in which we show that Roman-era-specific transport network connectivity does not influence interregional business link intensity in parts of Europe that were never part of the Roman Empire. Taken together, the exercises outlined above provide strong evidence that our estimates are, in fact, capturing Roman-era-specific effects.

Our paper relates and contributes to various literatures. Directly linked to our research is the literature concerned with identifying determinants of bilateral trade and especially the branch that assesses transport-cost related effects on trade flows (see, e.g., Duranton, Morrow and Turner, 2014; Pascali, 2017; Donaldson, 2018; Feyrer, 2019). We contribute by providing the first empirical evidence that transport costs, approximated by Roman effective distance, influenced trade during ancient (Roman) times. To the best of our knowledge, only the recent study by Barjamovic et al. (2019) applies a gravity-type framework to an earlier period (Bronze Age).

A related and rapidly growing literature investigates the contemporaneous and persistent effects of transport network accessibility on local economic activity (see, e.g., Michaels, 2008; Duranton and Turner, 2012; Jedwab and Moradi, 2016; Faber, 2014; Hornung, 2015; Redding and Turner, 2015; Donaldson and Hornbeck, 2016; Michaels and Rauch, 2016; Storeygard, 2016; Flückiger and Ludwig, 2019; Baum-Snow et al., 2017; Bakker et al., forthcoming). Particularly closely related to our paper are studies that specifically focus on the effects of Roman transport infrastructure. The recent paper by Dalgaard et al. (2018) documents that Roman road network density pre-determines modern road density and thereby influences the level of economic activity today. Similarly, Garcia-López, Holl and Viladecans-Marsal (2015), Percoco (2015) and De Benedictis, Licio and Pinna (2018) show that Roman roads influence current urbanisation patterns, firm locations, and transport costs via the routing of modern roads. Wahl (2017) shows that integration into the Roman Empire increases current-day economic activity. Again, persistence in access to the road network is identified as the main mediating factor. We complement these findings by considering all modes of transport in the Roman network—including waterborne transport—and documenting that, in addition to levels of development, historical connectivity influences the intensity of bilateral economic exchange. Although trade is very sensitive to shocks (Eaton et al., 2016), we show that the *relative* intensity in economic integration between regions is very stable in the long run.

Our study further informs the ongoing debate among historians of antiquity about whether or not Rome was a market economy. While there is broad consensus that staples, luxury goods, and a wide range of manufactured products were traded over long distances throughout the Ro-

man period (see, e.g., Hopkins, 1980; Horden and Purcell, 2000; Wilson and Bowman, 2018), the extent to which trade patterns were driven by market forces and trade costs rather than central planning remains debated (see, e.g., Whittaker, 1994; Polak, 2000; Mees, 2011; Willis, 2005; Fulford, 2018; Mees, 2018). We contribute to this discussion by providing first econometric evidence that the intensity of Roman trade in terra sigillata was indeed determined by transportation costs, suggesting that market forces mattered.[2] Importantly, our empirical approach enables us to circumvent issues related to preservation and excavation biases typically prevalent in the archaeological literature (Wilson, 2009).

Our findings directly speak to the literature on the determinants of interregional investment. Portes and Rey (2005) document that (geographical) distance deters exchange in financial assets. Similarly, Giroud (2013) and Campante and Yanagizawa-Drott (2018) show that airlink connectivity influences firms' decisions of where to invest. Leblang (2010) and Burchardi, Chaney and Hassan (2019) find that social ties created by historical migration are important determinants of foreign direct investment. They identify information asymmetries as an important underlying mechanism. Similarly, Guiso, Sapienza and Zingales (2009) show that genetic and somatic similarity affect bilateral trust, which, in turn, influences investment flows between countries.[3] We show that infrastructure investments of the distant past can lead to increased similarity in preferences and values and thereby foster investment flows. In this regard, our paper is also related to a literature concerned with explaining differences in economic preferences across space (Tabellini, 2008; Chen, 2013; Galor and Özak, 2016; Litina, 2016; Falk et al., 2018).

Also linked to our paper is the literature on the network structure of trade. The fact that networks influence international trade in differentiated products has been established both theoretically and empirically (e.g., Chaney, 2014; Garmendia et al., 2012; Combes, Lafourcade and Mayer, 2005; Rauch and Trindade, 2002; Rauch, 1999). In the spirit of this literature, we focus on a trade network that was established when the Roman transport network was created and show that it strongly and continuously influences interregional interaction.

Finally, we also connect to the discussion about the determinants of business cycle comovement (see, e.g., Burstein, Kurz and Tesar, 2008; Cravino and Levchenko, 2017). Our results highlight that events of the distant past can influence interregional transmission of economic shocks. In our case, the intensity of transmission is determined by connectivity within the Roman transport network.

The remainder of the paper is structured as follows: In Section 2, we provide background information on the creation of the Roman transport network along with qualitative evidence of its effect on contemporary trade; characteristics of the traded Roman terra sigillata are also described. The data is presented in Section 3. Section 4 describes our empirical framework. Regression results are discussed in Section 5; threats to identification are then addressed in Section 6. We investigate potential channels underlying our main results in Section 7, before

[2] To our knowledge, Kessler and Temin (2008) is the only study that provides econometric evidence for trade costs influencing economic integration during the Roman era. They show that Roman grain price differentials decline in distance (based on six price pairs).

[3] Ahern, Daminelli and Fracassi (2015) document a negative relationship between cultural distance and the volume of cross-border mergers.

concluding with Section 8.

2 Background

This section serves two purposes. First, it describes the evolution of the Roman transport network and outlines how it created a new pattern of cross-regional economic integration within the empire. Second, it illustrates why terra sigillata excavated at archaeological sites is well-suited to measure the intensity of interregional trade during the Roman era.

2.1 The Roman transport network and its effect on economic integration

At the time of maximum territorial expansion around 117 CE, the multi-modal Roman transport network consisted of approximately 80,000 km of paved roads, 25,000 km of inland waterways and a vast number of well-established shipping routes along the Mediterranean and Atlantic coasts (Chevallier, 1972; Scheidel, 2014). Starting with the connection of the capital Rome to regions on the Italian Peninsula, the (spatio-temporal) growth of the network had closely followed the territorial expansion of Rome. Once occupied, soldiers built roads connecting and cutting through the newly annexed regions in order to facilitate supply shipments and bringing in reinforcements. To minimise building cost and travel times for troops, Roman engineers designed roads to follow straight lines over long distances, thereby often ignoring local geographic and socio-economic conditions (Davies, 1998; Laurence, 2002).[4] Progress in civil engineering, such as the newly developed ability to construct permanent bridges, helped with the straight-line routing of roads. While the construction and design of roads was determined by military-strategic aims, they were subsequently used for commercial as well as private transport and communication (see, e.g., Temin, 2012, p. 223).

Roadworks followed clear and technologically novel standards, with surfaces consisting of several layers of sand, gravel, and rocks as well as drainage systems (Berechman, 2003). Combined with the construction of new road segments in core and peripheral regions, these technological advances greatly increased the freight-carrying capacity of the road network (Adams, 2012). The embedding of the road system into a unified legal framework constituted a further important Roman innovation that facilitated overland transport.[5] Among other things, this ensured that roads remained in good repair (Berechman, 2003).

Similar to terrestrial transport, capacities and organisation of waterborne transport substantially changed during Roman reign (see, e.g., Schmidts, 2011).[6] Along with the size of boats

[4] Illustrating that straightness of routing was prioritised over ease of travel is the fact that many road sections did not meander and had steep gradients (Davies, 1998). The military-strategy and straight-line-preference-influenced routing of roads further suggests that roads were not systematically built to connect existing settlements (see, e.g., Laurence, 2002).

[5] Roads were categorised into four groups (*via publica, via militaris, via vicinalis, via privata*), with functions and entities responsible for maintenance clearly defined (Rathmann, 2003). *Viae publicae*, for example, were constructed and maintained by the state. Maximum load allowances for carts reduced the wear and tear of the pavement (Berechman, 2003). Roads were required to support the heaviest category, i.e. carts up to a weight of 1,500 Roman pounds (around 500 kg) drawn by two pairs of oxen.

[6] An example of institutional change is that river transport was to a large extent controlled by well-organised

and ships, the quantity of goods shipped via waterways increased dramatically. Large flat-bottomed barges used for river transport were able to carry around 150 tonnes of cargo. Seagoing ships were even loaded with up to 1,000 tonnes of freight (Campbell, 2012, p. 217). Canals—typically constructed to bypass dangerous parts of rivers or to facilitate navigation through river deltas—also contributed to the reduction of water transportation costs (McWhirr, 2002). Adding to the innovations in terrestrial and waterborne transport infrastructure, the empire-wide (political) stability and peace (pax Romana) further stimulated the establishment and deepening of long distance trade relationships (Sidebotham, 1986, p. 181). Piracy in the Mediterranean, for example, a previously common and trade-deterring problem, was largely suppressed after 67 BCE (de Souza, 2002, p. 96). The introduction of a common currency as well as improvements in container technologies (amphorae and barrels) further facilitated long distance trade (see Wilson and Bowman, 2018, p. 5–6).

Information on cross-regional economic interaction before Roman occupation is scarce.[7] While certainly existing, trade among tribes or between Roman merchants and tribes was comparatively limited and localised prior to occupation. The amount of Roman goods excavated in Celtic regions (such as amphorae and other pottery products) that pre-date the occupation is low (Fitzpatrick, 1985, p. 310). Following annexation and integration into the empire-wide transport network, diversity and quantity of goods exchanged substantially changed. Agricultural surpluses of the former Celtic and Egyptian regions, for example, crucially contributed to the food security of Rome (Erdkamp, 2013). Similarly, new types of cereals were imported from southern provinces (Reddé, 2018, p. 147). Access to the transport network also promoted specialisation and the exchange of manufactured products. Various commodities—e.g. amphorae, ceramics, glass, lamps, bronze statuettes—were produced in large quantities at centralized production sites and traded over long distances (Bowman and Wilson, 2009, p. 17). Accompanying economic interaction, the transport network increased interpersonal interaction and induced migration as well as technological and cultural diffusion (see, e.g., Willis, 2005). For example, the custom of sharing meals was spread by Roman soldiers (Willis, 2005, ch. 7.2.2). As a result of such exchange, similar goods and technologies could be found across all Roman provinces (see Wilson and Bowman, 2018, p. 5).[8] The 'Roman consumption package' consisting of amphorae for wine, olive oil, fish products, and table pottery was available throughout the empire (Bowman and Wilson, 2009, p. 17).

In sum, the Roman Empire-wide integrated transport network led to an unparalleled degree of market integration and created a new pattern of interregional (socio)-economic exchange (Bowman and Wilson, 2009, p. 17). While pre-existing roads and waterways may have facilitated initial Roman occupation, the 'barbarian regions' had not been part of an *integrated*

cooperations of *nautae* (boatmen) (Schmidts, 2011).

[7] For Celtic Gauls there is evidence of considerable trading activity. Ships, for example, were used for river transport. Furthermore, they maintained ports in Britain to control trade with this region. Shipwrecks discovered in the Mediterranean additionally hint at a Celtic ship-building tradition (Schmidts, 2011, p. 93).

[8] Hitchner (2003, p. 398) emphasises that "A citizen of the empire travelling from Britain to the Euphrates in the mid-second century CE would have found in virtually every town along the journey foods, goods, landscapes, buildings, institutions, laws, entertainment, and sacred elements not dissimilar to those in his own community."

supra-regional transport network.[9] Furthermore, technology, routing, density, and maintenance of transport infrastructure substantially changed after Roman annexation. These alterations, along with the unprecedented geographical reach of the network imply that the (bilateral) accessibility between regions dramatically changed (e.g., Hitchner, 2012). We provide empirical support for this notion in Section 6.

The cost of shipping goods between regions potentially plays a dominant role in explaining how the Roman transport network shaped the pattern of bilateral exchange. Although disputed among early historians of antiquity (see, e.g., Finley, 1999; Jones, 1964; Yeo, 1946), it is plausible that the intensity of trade between regions depended on the costs of transportation. These were influenced by the available means of transport and their associated per unit freight rates. The latter varied dramatically across modes and reflected efficiency differences between Roman transport technologies. On the basis of emperor Diocletian's *Edict on Maximum Prices* from 301 CE—an original contemporary source—Scheidel (2014) recently revised existing estimates of relative per-unit-distance transport costs (see Appendix A.4 for more details). They show that seaborne transport was the most cost effective mode of shipping with a (normalised) per unit distance freight rate of one, followed by downstream and upstream river transport with associated costs of 5 and 10, respectively. Road transport was by far the most expensive way of moving goods. The historical freight rate data suggest a cost of 52 relative to seaborne transport.[10] Qualitative accounts and case studies indicate that these transport-mode-dependent cost differentials influenced the decision along which routes to ship goods. The geographical distribution of archaeological pottery finds produced at Banassac in the south of France, for example, implies that indirect routes were chosen over distance- or time-wise shortest paths in order to make use of cost-effective means of transport, i.e., sea or river (Mees, 2011, p. 260).

To date, there is no systematic assessment of the effects of transport costs on interregional trade during the Roman era. The first principal aim of this paper is to fill this research gap. To this end, we require historical data on bilateral transport costs and trade volumes. The former can be inferred from the structure of the Roman transport network and relative freight rate differentials across shipping modes. As outlined below, the spatial distribution of terra sigillata excavations allows for the reconstruction of trade flows.

The second aim of this paper is to analyse how differences in connectivity within the Roman transport network influence economic integration today. In this context, it is important to note that today's routing of roads is strongly influenced by the paths chosen by the Roman engineers.[11] Furthermore, relative transport costs across shipping modes were relatively stable

[9] Until the defeat of Vercingetorix by Caesar in 46 BCE, for example, Romans used local Gaulish roads and seized Gaulish ships to move troops (Chevallier, 1972, p. 25). However, since Gaulish tribes were not unified, no coherent concept of road building, let alone an integrated cross-regional transport network designed for purposes of trade or military campaigns, existed.

[10] While there is some debate about the appropriate estimates of *absolute* levels of transport costs among historians, there is broader consensus that the above-mentioned cost ratios are reflective of *relative* freight rate differentials during Roman times (see Scheidel, 2014, p. 9). The first price-edict-based estimates produced by Duncan-Jones (1974), for example, suggest the following cost ratios: 1 (sea), 4.9 (river), and 34–42 (road). Additionally taking differences in upstream and downstream river transport into account, more recent studies estimate relative costs of 1 (sea), 5 (downriver), 10 (upriver), 34–42 (road) on the basis of the price edict (Franconi, 2014, p. 57).

[11] There are many examples of today's highways following Roman roads. Well-known stretches include Arles to

until the advent of the transport in the 19th century (c.f. Johnson and Koyama, 2017; Masschaele, 1993). However, with the introduction of new transport technologies, such as steam engines, railways and later on aeroplanes, cost ratios changed substantially (e.g., Feyrer, 2019; Donaldson, 2018).

2.2 Production and trade of terra sigillata

Gallo-Roman terra sigillata is a red-gloss tableware made out of clay which was manufactured at several large production centres in Italy (est. 1st century BCE), Gaul (est. 1–2 century CE), Germania and Raetia (est. 2–3 century CE). These centres, whose location were determined by clay deposits, produced millions of pieces using an unprecedented division of labour. At La Graufesenque (South France), for example, batches of more than 30,000 vessels were common; kiln firings reached very high temperatures (around 950 degrees Celsius) and were shared by up to twelve potters (Marichal, 1988; Polak, 1998).[12] Potters stamped their names in the inside of vessels to identify their works and distinguish between production batches (Wilson, 2009, p. 397). Based on these stamps, each piece of tableware can be traced from production site to the location of consumption, where it was later excavated by archaeologists. This ability to identify origin and destination of (stamped) products is—in the context of our study—a core property of terra sigillata.

A second aspect that makes it well-suited for our analysis is its widespread use. Measured as a share of Roman trade, terra sigillata accounted for approximately 10 percent of total volume and an even higher proportion of value (Mees, 2018).[13] High-quality Gallo-Roman terra sigillata—often produced at kiln sites located in hard-to-reach inland regions—was traded across most of the Mediterranean, the Northwestern Empire, the Danube region, and the Barbaricum up to Poland. Low quality ceramic cooking ware and amphorae, in contrast, were almost exclusively manufactured at coastal kiln sites, thus allowing for a cost-effective distribution (see Wilson and Bowman, 2018, p. 10–11). Due to the wide range of terra sigillata products—such as bowls, cups, platters, amphorae, and mortaria—demand stemmed from a great variety of sources, including public, private and commercial entities located in urban as well as rural areas. The distribution of terra sigillata was organised in sophisticated logistics chains. Rather than directly delivered to individual costumers, it was typically shipped in bulk from production sites to warehouses and shops (Willis, 2005, ch. 6.4.6). Terra sigillata produced at La Graufesenque and destined for consumption in the northern border region of the empire, for example, was first transported via mountainous roads to Narbonne. There, it was transferred to barges and shipped upstream on the Rhône to Lyon, the regional trade centre. It was then stored in warehouses until further distribution (Mees, 2011).

Aix, Clermont-Ferrand to Limoges, Arcachon to Bordeaux, Saintes to Poitiers. In fact, the surfaces of these roads consisted of the original Roman cobbles and gravel until the introduction of railways in France (Hitchner, 2012). Likewise, British Ordnance Survey Maps document that approximately 3,200 km of modern roads follow Roman road trajectories. Three of the four royal highways of medieval Britain were originally built by the Romans.

[12] Figure A.3 depicts (examples of) kiln sites and excavated terra sigillata products.

[13] The price for a piece of terra sigillata typically ranged from 12 to 20 *asses*, equivalent to the daily pay of a soldier (Darling, 1998, p. 169).

The geographical distribution of production and excavation sites of stamped terra sigillata—on the basis of which we construct our measure of bilateral trade intensity—is depicted in Figure 1.Possibly important factors explaining the varying penetration of different terra sigillata products are taste for variety, variation in quality, and shipping costs. Depending on the available transport modes, the latter could vary greatly, even for two regions located equidistant from a given production site. By employing a gravity-type estimation approach, we isolate the effect of transport costs from other factors and estimate to what extent they influenced interregional trade flows and thus help explain the spatial distribution of archaeological finds.

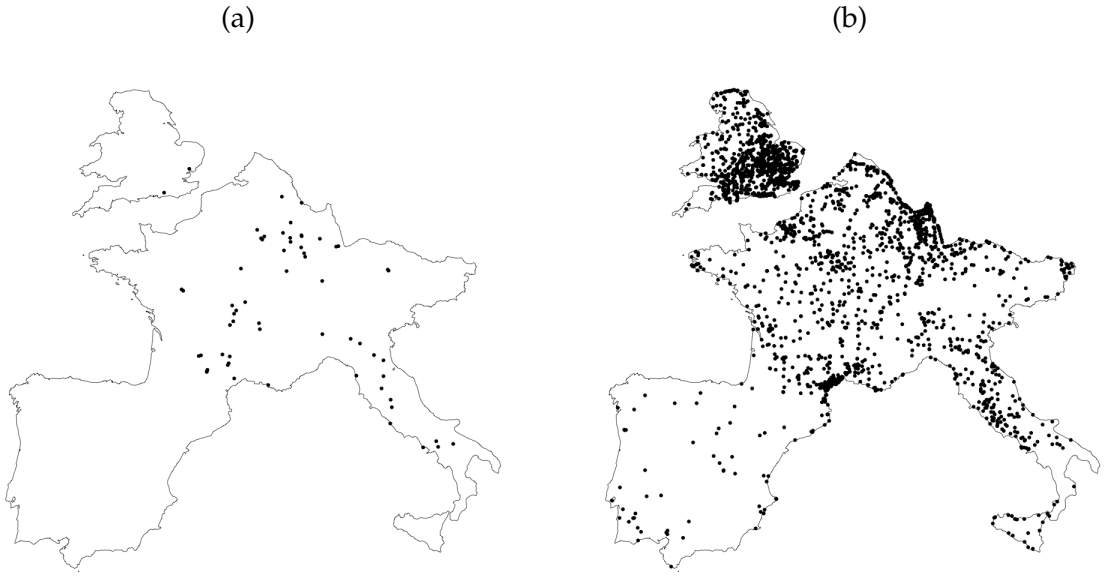


Figure 1: Origins and Destinations of Roman Terra Sigillata
Panel (a) depicts the locations of terra sigillata production sites; panel (b) shows the spatial distribution of terra sigillata excavation sites. The figure is restricted to the geographical scope of our analysis (see Section 3).

3 Data

Our study covers the regions within Austria, Belgium, England, France, Germany, Italy, Luxembourg, Netherlands, Portugal, Spain, and Switzerland that were once part of the Roman Empire. For the empirical analysis, we divide this area—referred to as 'Western Europe'—into grid cells of 0.5×0.5 degrees longitude/latitude (ca. 55×55 kilometres). In our main analysis, we only consider cells that are intersected by the Roman transport network. Illustrating the high density of the network, the 903 intersected grid cells cover 88% of the territory of Western Europe.[14] Based

[14]Figure A.1 depicts the grid cells that are intersected by the Roman transport network. In Tables C.1–C.2, we show that our findings remain unchanged when we incorporate non-intersected cells into the network by creating artificial road connections. The motivation for excluding the non-intersected cells in our main analysis is that they are structurally different because their 'artificial connection' from the centroid to the network is substantially longer compared to intersected cells (24.2 vs 7.5 kilometres). See below for more details.

on the 903 cells we construct grid-cell-pair-level measures of (i) transport network connectivity, (ii) economic integration during Roman times, and (iii) current-day intensity of economic ties.

3.1 Transport network connectivity

We predict the cost of transporting goods between two regions during the Roman era under the assumption that agents can use the full, empire-wide, Roman transport network at its maximum extent (reached in year 117 CE). To this end, we combine information on location of Roman roads, navigable rivers, and coastal routes. The road network is extracted from the digitised version of the Barrington Atlas of the Greek and Roman World (Talbert and Bagnall, 2000). We identify navigable river sections that the Romans used for transport using a wide range of historical sources (listed in Table A.2). Transport by sea is possible along the coast. Combined, roads, navigable rivers, and coastal routes, make up our multi-modal Roman transport network.[15] This network—depicted in Figure 2—is subsequently denoted by $\mathbf{N}^{Rome}$ and represents a collection of numerous segments which differ in length and associated mode of transport. As outlined above, the cost of shipping goods over a given distance varied substantially across transport modes. These Roman technology-driven differences in relative shipping costs are captured by the vector $\boldsymbol{\alpha}^{Rome} \equiv \left(\alpha^{sea}, \alpha^{river}, \alpha^{road}\right)$. We normalise freight rates such that $\alpha^{sea} = 1$; drawing on Scheidel (2014) we set $\boldsymbol{\alpha}^{Rome} = (1,\ 7.5,\ 52)$. The relative cost of shipping goods via rivers (7.5) represents the average between up- and downstream freight rates (5 and 10, respectively).[16]

To predict transport costs between two grid cells, we assume that agents choose the cheapest among all possible routes given the Roman-specific transport cost differentials $\boldsymbol{\alpha}^{Rome}$ and transport network $\mathbf{N}^{Rome}$.[17] The least-cost path is identified using Dijkstra (1959)'s algorithm, where the geographical centres (centroids) of grid cells are set as origins and destinations.[18] Throughout, we assume that transshipment between different transport modes is costless. Following Donaldson (2018), we refer to the costs associated with transporting goods along the optimal path as the 'Roman effective distance'. Subsequently, this cost is denoted by $LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})$ and we employ the natural logarithm of this measure as our main explanatory variable. It is important to note that variation in Roman effective distance is generated by the

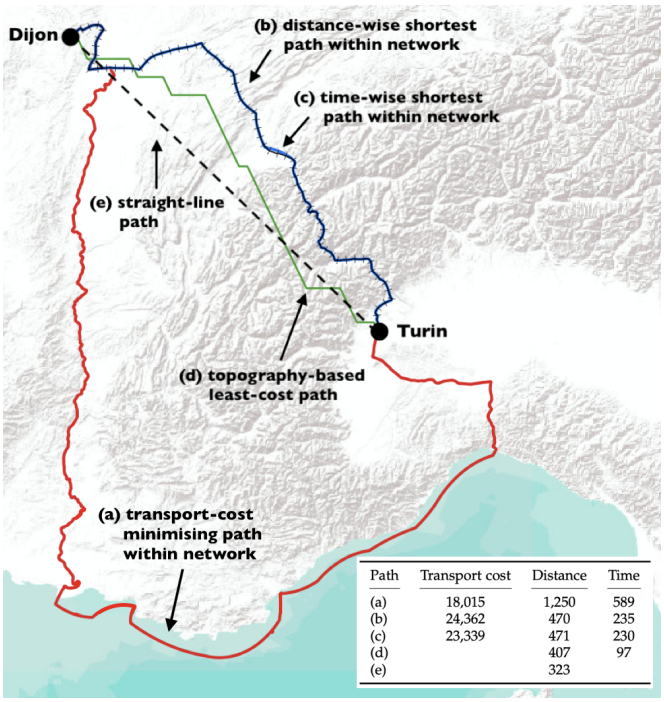
[15] Compared to the Stanford geospatial network model of the Roman world (ORBIS) our data source offers a greater geographical coverage in terms of routes and sites. Furthermore, the broad spectrum of information that is used by ORBIS to compute transport costs raises concerns that connectivity within the ORBIS transport network is partly determined by observed interaction (i.e., endogenous) during Roman times. Network segments, for example, are ranked according to their significance.

[16] We use the undirected rather than the directed transport network to identify the least cost paths. Two reasons motivate this choice. First, when analysing the effects of the Roman transport network on the intensity of interregional business links today, it is not *a priori* clear how transport-direction-dependent cost differentials should affect the direction of investment flows. Second, in auxiliary regressions discussed in Section 6, we employ measures of bilateral interaction that do not allow for a distinction between origin and destination. In these cases, we would have to arbitrarily impose a directed structure. As illustrated in Tables C.1–C.2, results are similar if we use the directed instead of undirected transport network to predict shipping costs.

[17] Agents are allowed to use segments of the Roman transport network that lie outside Western Europe.

[18] Similar to Donaldson and Hornbeck (2016), grid cell centres are connected to the transport network by creating an artificial straight-line road segment between the centroid and the closest point on the section of the network that intersects the grid cell. This procedure is illustrated in Appendix A.3. On average, we create an artificial road of 7.5 kilometres in length, representing 5.7% of the total cost of the optimal path. Results remain stable if we vary per-unit costs of the artificial connections within the range zero to 200. Estimates are presented in Table C.3.

Figure 2: The Multi-Modal Roman Transport Network
Map shows the Roman transport network (restricted to the geographical scope of our analysis). Grey lines symbolise roads, solid black lines navigable river sections, and dashed lines coastal shipping routes.



Path	Transport cost	Distance	Time
(a)	18,015	1,250	589
(b)	24,362	470	235
(c)	23,339	471	230
(d)		407	97
(e)		323	

Figure 3: Least-Cost Paths
Map depicts five different least-cost paths between Turin and Dijon: (a) The least-cost path within the Roman transport network, given $\mathbf{N}^{Rome}$ and Roman-specific technology $\boldsymbol{\alpha}^{Rome}$ (red). (b) The distance-wise shortest path within the Roman transport network (blue). (c) The time-wise shortest path within the Roman transport network (black cross-lined). (d) The topography-based least-cost path identified using the Human Mobility Index with Seafaring (Özak, 2018) (green). (e) The straight-line (as the crow flies) path. The length of this path is equal to the geodesic distance (dashed black). 'Transport cost' refers to the cost of shipping goods along a given path (i.e., the freight-rate-weighted distance). 'Distance' refers to the distance of a given path (measured in kilometres). 'Time' refers to the shipping time along a given path (measured in hours).

combination of the network structure ($\mathbf{N}^{Rome}$) and freight-rate differentials ($\boldsymbol{\alpha}^{Rome}$).

To gain an intuitive picture of the difference between the transport-cost minimising and other least-cost paths, Figure 3 depicts three separate types of least cost paths *within* the Roman transport network that connect Turin to Dijon: (a) The transport-cost minimising path within the Roman transport network given the Roman transport technology. Costs associated with shipping goods along this path are referred to as Roman effective distance (i.e., $LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})$). (b) The distance-wise shortest path within the Roman transport network. The costs associated with using this path—which we subsequently refer to as network distance—are equal to the length of the path (measured in kilometres). (c) The time-wise shortest path within the Roman transport network, where costs are expressed in hours (referred to as network time).[19] Additionally, Figure 3 depicts two commonly used least-cost paths that are independent of the structure of the Roman transport network: (d) The topography-based least-cost path identified on the basis of the Human Mobility Index with Seafaring (HMISea, Özak, 2018). This index is a proxy for pre-industrial human mobility and takes into account human biological constraints as well as geographical and technological factors. The HMISea least-cost path is not dependent on the transport network structure $\mathbf{N}^{Rome}$. The costs associated with this optimal path are captured by travel time (in hours). (e) The straight-line path (as the crow flies). The length of this line—also interpretable as costs—is equal to the geodesic distance between Turin and Dijon.

Figure 3 visualises two important points: First, within the Roman transport network, the path that minimises transport costs differs markedly from the distance- as well as the time-wise shortest paths. The differences are exclusively driven by the mode-dependent shipping costs ($\boldsymbol{\alpha}^{Rome}$). When seeking to minimise transport costs, there is a trade-off between (i) minimising distance covered when transporting goods between two locations and (ii) reaching and making use of cost-effective modes of transport (i.e., minimising average transport costs per kilometre). In the example of Figure 3, this trade-off results in a substantial detour of the transport-cost minimising path. The table in Figure 3 illustrates the effects of $\boldsymbol{\alpha}^{Rome}$. The transport-cost minimising path is more than 2.7 times longer compared to the shortest possible route and 2.6 times slower than the fastest route. However, because the detour allows for the use of more cost-effective means of transport, overall shipping costs are more than 20% lower compared to transport over the distance- or time-minimising path.[20]

The second important point illustrated in Figure 3 is that the distance- and time-wise shortest paths are almost identical (apart from a segment that crosses the Lake Geneva) and connect Dijon to Turin in a relatively direct line. Similarly, the topography-based least-cost path (HMISea) does not take any major detours. This suggests that the cost of transporting along these three optimal paths proportionally increases with geodesic distance. In contrast, the non-linearity of the transport-cost minimising path indicates that the correlation between the Roman

[19] To identify the time-wise shortest path, we combine the network $\mathbf{N}^{Rome}$ with mode-specific travel speeds $\boldsymbol{\alpha}^{\text{Time}} = (3.7,\ 1.565,\ 2)$ from Carreras and Soto (2013). Differences in travel speeds, measured in km/h are relatively small.

[20] The table also shows that the quickest path within the network takes more time than the geography-based least-cost path (HMISea). Two factors explain this difference. First, the HMISea captures the time it takes humans to move between location whereas the time-wise shortest path within the Roman transport network specifically captures shipping time of goods. Second, movement is not restricted to follow the network in the case of the HMI.

effective distance and geodesic distance is limited.

Table 1 confirms that this conjecture is borne out in the data. Column (1) shows that within our historical sample (see details below), the correlation between Roman effective distance and geodesic distance is 0.38. That is, Roman effective distance does increase in geodesic distance, but this effect is limited. The correlation with the remaining three least-cost measures is similar in magnitude, ranging from 0.36–0.47.[21] Figure A.4 in Appendix A.8 illustrates that these relatively low correlations are not driven by a specific part of the Roman effective distance distribution.

Costs of the distance- and time-wise shortest paths within the Roman network, on the other hand, are extremely highly correlated with geodesic distance (column 2). Correlation is also high between geodesic distance and the topography-based HMISea measure. This implies that the variation in these three least-cost measures is largely captured by geodesic distance. On the other hand, a large part of the variation in Roman effective distance is specifically due to combination of the layout of the Roman transport network ($\mathbf{N}^{Rome}$) and transport technology ($\boldsymbol{\alpha}^{Rome}$). To isolate the Roman-era specific aspects in the subsequent empirical analysis, we account for geodesic and other topography-based distances.

Table 1: Bivariate correlation coefficients between least cost measures

	ln Roman effective distance (1)	ln geodesic distance (2)	ln network distance (3)	ln network time (4)
ln geodesic distance	0.379			
ln network distance	0.393	0.982		
ln network time	0.468	0.976	0.984	
ln HMISea	0.357	0.934	0.915	0.904

Notes: This table presents bivariate correlation coefficients between the least-cost measures depicted in Figure 3, based on the historical sample used in Table 2. 'effective distance' represents the cost associated with shipping goods along the least cost path between grid cells, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'geodesic distance' represents the length in kilometres of the straight-line path (as the crow flies) between grid cells. 'network distance' represents the length in kilometres of the distance-wise shortest path between grid cells, given the Roman transport network. 'network time' represents the travel time in hours along the time-wise shortest path between grid cells, given the Roman transport network and Roman-era-specific speed for each mode of transport. 'HMISea' represents the travel time in hours along the topography-based shortest path between grid cells, identified based on the methodology in (Özak, 2018) that incorporates only geographical features and pre-industrial technology.

3.2 Measuring economic integration during the Roman era: terra sigillata

To measure bilateral trade volumes during Roman times, we extract information on terra sigillata finds from the Roman tableware database which has recently been made available online by the Römisch-Germanisches Zentralmuseum in Mainz.[22] The stamped vessels were produced between the beginning of the first century and the middle of the third century. Between 75–125 CE a range of terra sigillata products were not stamped. These unstamped items amount to approximately 30% of total excavated terra sigillata (Furger and Deschler-Erb, 1992, Fig. 84).[23]

[21] Correlation coefficients are very similar—and even somewhat lower—in our current-day sample (Table A.4).

[22] www.rgzm.de/samian. The samian data is based on the publications of Names on Terra Sigillata (see Hartley et al., 2008) and the Corpus Vasorum Arretinorum (see Oxé, Comfort and Kenrick, 2000).

[23] This implies that terra-sigillata-based estimates of variation in trade over time would need to be interpreted with caution and may suffer from measurement error.

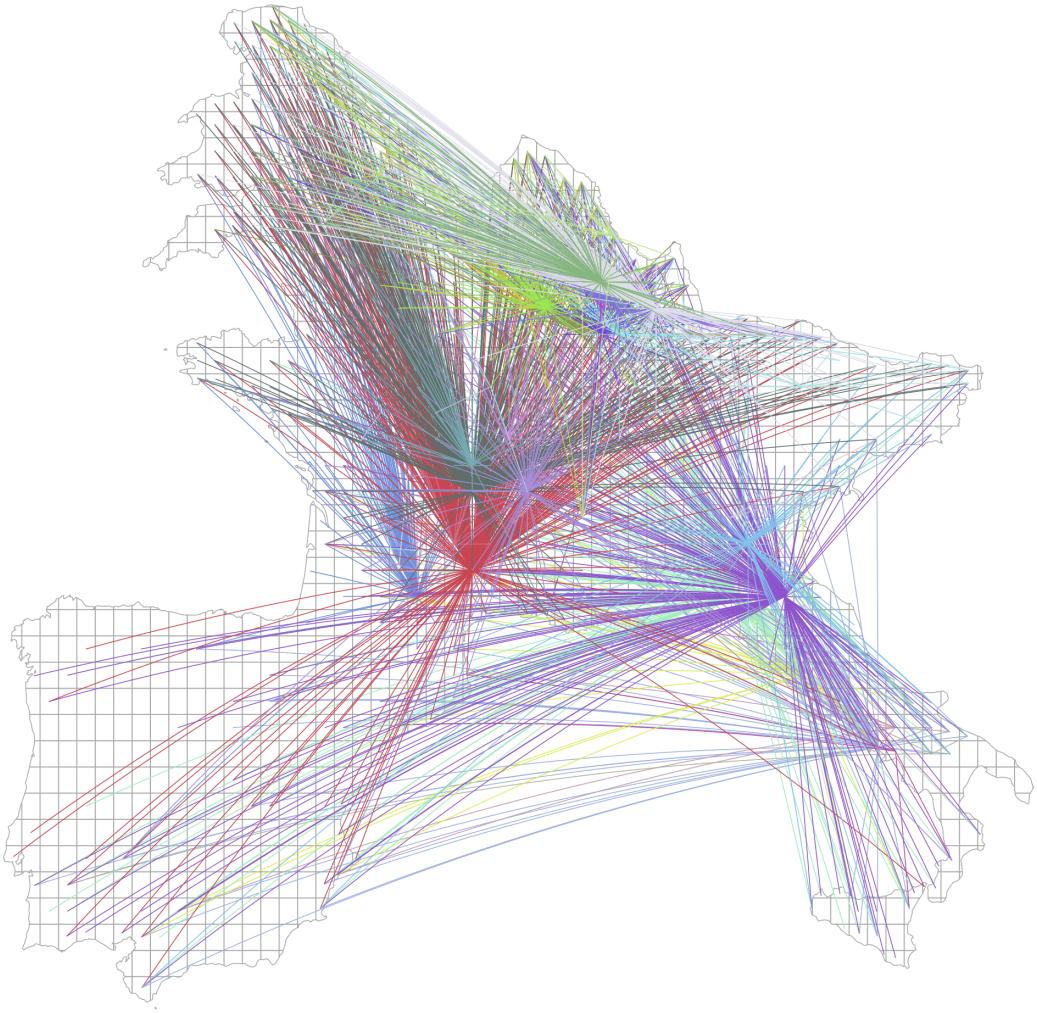


Figure 4: Bilateral Trade Flows in Terra Sigillata during Roman Era
Figure maps trade flow shares of terra sigillata between grid-cell pairs. Each colour is specific to one origin grid cell.

Crucial for our analysis, there is no indication that shipment and distribution of these types systematically differed from stamped types.

Based on the precise information on the site of production, identified via the potters' stamp, and location of excavation, we assign each find to its grid cell of origin and destination. We then aggregate this information to the grid-cell-pair level giving us the number of terra sigillata finds within grid cell j that were produced in grid cell i.[24] Following Eaton, Kortum and Sotelo (2013), we define the share of j's total number of ceramics that originate from i as our

[24]This variable therefore captures trade volumes rather than values.

measure of interregional trade flows.[25] The 56 individual production sites fall into 44 different grid cells. For the Roman era, we thus have 44 origin grid cells from which goods can potentially be shipped to the 903 grid cells that are connected to the network. For 520 of these grid cells, we observe at least one terra sigillata find manufactured in any one of the 44 'production grid cells'. Because we employ a Poisson pseudo-maximum-likelihood estimator and account for origin and destination fixed effects in our data analysis, any grid cell with zero terra sigillata finds *and* zero production sites is excluded due to collinearity (see Appendix A.1 for more details). Abstracting from within grid-cell trade, our dataset for the historical analysis consists of 22,839 observations. Figure 4 visualises the trade flows (expressed in shares of total imports); Table A.5 reports summary statistics of the key variables.

3.3 Measuring economic integration today: cross-regional firm ownership

The number of cross-regional firm ownership links is based on the Bureau van Dijk's Orbis database. This database covers around 300 million companies worldwide and contains detailed firm-level information on industry, location, and ownership. For our analysis, we focus on firms with an annual operating revenue of more than 2 million U.S. dollars. The data was downloaded between February–April 2018, and consequently captures a snapshot of ownership patterns at that point in time. To compute the grid-cell-pair number of business links, we first identify all firms that are located within Western Europe. Among these firms we then extract the subset of companies that are in a cross-regional parent-subsidiary relationship. Specifically, we keep all firms that either own a stake of at least 25% in another firm that is domiciled in a different grid cell or that are 25% owned by a company registered in another grid cell. The location of these firms was geocoded manually. For our analysis, we are left with 106,996 cross-regional parent-subsidiary links. These business links are aggregated to the grid-cell-pair level by counting the number of firms located in 'destination grid cell' j that are (part-)owned by firms registered in 'origin grid cell' i. Again, we use shares—i.e., bilateral inflow divided by total inflows—as measure of interaction intensity. Our final grid-cell-pair-level dataset consists of 731,823 observations, made up of 865 origin and 847 destination grid cells for which we observe at least one non-zero investment flow. For summary statistics see Appendix A.1.

In extensions to our main analysis, we make use of further data sources (price correlations, onset of the Black Death, and night-time light intensities). These data are described as they become relevant (see also Appendices B–I).

4 Empirical framework

To explain the bilateral (socio-)economic integration—past and present—we rely on the gravity framework. The gravity framework has many micro-economic foundations (see Yotov et al., 2016,

[25] For the main analysis, we aggregate trade flows across production sites within grid cells. Our results remain qualitatively unchanged if we aggregate trade flows to the production-site level and run regressions at the production-site-destination-grid-cell level. To that end, we augmented Eq. (1) to include production site and destination fixed effects. The results are presented in Table C.4.

for a discussion). The underlying data for our dependent variable is, similar to Barjamovic et al. (2019), a count of finds rather than a volume of trade. Hence, we base our specification on the finite-sample version of the gravity framework developed by Eaton, Kortum and Sotelo (2013) and use shares rather than counts as outcome. We therefore estimate the following regression model using the Poisson Pseudo Maximum Likelihood (PPML) estimator:[26]

$$\mathcal{X}_{ij} = \exp\left(\delta \ln LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij} + \theta\, d_{ij} + T'_{ij}\,\gamma + \beta_i + \beta_j\right) + \varepsilon_{ij}, \tag{1}$$

where $\mathcal{X}_{ij}$ denotes the share of imports in grid cell j that originate from grid cell i, i.e., the number of pottery finds in j originating from i relative to all pottery finds in j. The main explanatory variable is the least-cost path effective distance, $LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij}$. The coefficient δ captures the elasticity of economic integration with respect to the Roman effective distance. As discussed above, variation in Roman effective distance is generated by both the structure of the transport network $\left(\mathbf{N}^{Rome}\right)$ and the mode-specific differences in transport costs $\left(\boldsymbol{\alpha}^{Rome}\right)$. To isolate the Roman-era specific part of the variation, we condition on geodesic distance (d_{ij}) as well as geographical and historical factors. The latter are subsumed in the vector T_{ij}.

Throughout our analysis, we control for the full set of origin and destination fixed effects (represented by β_i and β_j, respectively). These dummies control for market size which, in addition to trade costs, is a central feature of gravity-type equations. They also absorb any other differences in region-specific characteristics—such as income levels or geographical location—that influence the overall level of economic integration. In the context of archaeological data it is important to note that the fixed effects wash out potentially existing excavation biases, i.e., the possibility that discovering Roman tableware is more likely in economically more integrated and populated areas. Finally, the inclusion of origin dummies also controls for production-site-specific quality differences that influence the magnitude of interregional trade flows.[27] The error terms ε_{ij} are clustered along two dimensions: the origin and destination grid-cell level.

The crucial assumption underlying our estimation strategy is that conditional on control variables, Roman effective distance captures Roman-era-specific variation in transport network connectivity and is uncorrelated with factors in the error term that influence outcomes. There are two primary threats to the validity of this assumption. The first is that Roman effective distance is endogenous to pre-existing patterns of interaction. For example, it is possible that roads were built to more directly connect regions that already interacted more intensely.[28] The second is

[26] As shown by Sotelo (2019), estimating the gravity model with the Multinomial Pseudo Maximum Likelihood (MPML) estimator developed by Eaton, Kortum and Sotelo (2013) produces the same estimates as the Poisson Pseudo Maximum Likelihood (PPML) estimator from Santos Silva and Tenreyro (2006) when destination fixed effects are included. As we have comparably many fixed effects, estimates are performed with the Stata command `ppmlhdfe` developed by Correia, Guimarães and Zylkin (2020).

[27] Baldwin and Harrigan (2011) discuss the possibility of identifying valid trade theories (including theories that do account for quality differences) by looking at quantities, values, and prices. Lacking information on the latter two dimensions, we cannot identify which theory most accurately explains trade flows during Roman times. However, as we are interested in investigating the effect of trade costs on the bilateral allocation rather than assessing the validity of specific theories, controlling for quality differences using origin and destination fixed effects is sufficient.

[28] Note that endogeneity in placement of roads does not constitute a threat to identification in the context of our analysis. The fact that a grid cell is cross-cut by (multiple) roads is absorbed by the origin and destination fixed effects. Furthermore, we only include grid cells that are intersected by at least one segment of the Roman transport network.

that Roman effective distance partially captures geographic connectivity.

In Section 6, we address these concerns and document that connectivity within the Roman transport network does not predict intensity of interregional interaction in pre-Roman times. Furthermore, we develop an instrumental variable strategy to mitigate the concern that routing captures geographic features. The approach exploits the fact that the routing of Roman roads was primarily influenced by military-strategic considerations. We additionally show that our estimates remain stable when we account for a variety of geography-based least-cost measures as well as a wide range of geographical aspects. Finally, we conduct a falsification exercise in which we document that Roman-era-specific transport network connectivity does not influence interregional business link intensity in regions of Europe that were never part of the Roman Empire.

5 Main results

In this section, we first document that effective distance within the Roman transport network determined the geography of Roman trade. In the second step, we move to the current day and show that variation in Roman transport network connectivity is reflected in today's spatial firm ownership structure. Possible threats to the validity of our estimation strategy are discussed in detail in Section 6.

5.1 Roman transport network connectivity and economic integration during the Roman era

We start our empirical analysis by regressing bilateral terra sigillata trade shares on Roman effective distance as well as origin and destination fixed effects. The point estimate obtained from running this parsimonious specification can be directly interpreted as the elasticity of trade with respect to distance. Column (1) of Table 2 documents that transport costs strongly deter Roman trade. The statistically highly significant point estimate of -2.895 implies that a one percent increase in Roman effective distance reduced bilateral trade by almost 3%.

In column (2), we control for the number of centuries that two grid cells jointly spent under Roman rule. This variable accounts for the fact that total trade volumes potentially increase with time belonging to the same economic and political entity. The fact that two regions were connected through the Roman transport network only once both had become part of the empire increases this likelihood further. Confirming expectations, we observe that the total trade volume between two regions increases by 294% with each additional century shared under Roman reign.[29],[30] Compared to column (1), the Roman effective distance coefficient decreases by

In the context of our study, issues arise only if bilateral (i.e. grid-cell-pair-specific) aspects systematically influenced the routing of the network.

[29]Note that the PPML estimator specifies the conditional mean as $E[\mathcal{X}_{ij}|\mathbf{X}] = \exp(\mathbf{X}\boldsymbol{\beta})$, where $\mathbf{X}$ collects all explanatory variables. Hence, the marginal effect of the exogenous variable x_k is given by $\partial E[\mathcal{X}_{ij}|\mathbf{X}]/\partial x_k = \exp(\mathbf{X}\boldsymbol{\beta})\beta_k$. Reformulating leads to $(\partial E[\mathcal{X}_{ij}|\mathbf{X}]/\exp(\mathbf{X}\boldsymbol{\beta}))/\partial x_k = (\partial E[\mathcal{X}_{ij}|\mathbf{X}]/E[\mathcal{X}_{ij}|\mathbf{X}])/\partial x_k = \beta_k$, which implies that the coefficients can be interpreted as semi-elasticities.

[30]Due to the fact that we do not have detailed information on timing, neither on trade flows nor on the evolution of

Table 2: Roman transport network connectivity and trade during the Roman era

Dependent Variable:	Share of Terra Sigillata Finds					
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	-2.895***	-2.100***	-2.053***	-1.493***	-1.498***	-1.254***
	(0.593)	(0.559)	(0.509)	(0.542)	(0.516)	(0.467)
Joint duration under		2.943***	2.638***	2.277***	2.277***	2.287***
Roman rule (centuries)		(0.294)	(0.379)	(0.277)	(0.277)	(0.263)
ln geodesic distance				-0.655***	-0.679	-0.038
				(0.230)	(0.674)	(0.554)
ln network distance					0.029	
					(0.680)	
ln network time						-0.886
						(0.601)
Geography controls	No	No	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,839	22,839	22,839	22,839	22,839	22,839
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid-cell level are reported in parentheses. Dependent variable is the share of terra sigillata finds in cell j that originates from cell i. 'Roman effective distance' represents the cost associated with shipping goods along the least cost path between grid cells, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'Joint duration under Roman rule' is the number of centuries two grid cells were jointly under Roman rule. 'geodesic distance' represents the length in kilometres of the straight-line (as the crow flies) between grid cells. 'network distance' represents the length in kilometres of the distance-wise shortest path between grid cells, given the Roman transport network. 'network time' represents the travel time in hours along the time-wise shortest path between grid cells, given the Roman transport network and Roman-era-specific speed for each mode of transport. Geography controls include the absolute difference in latitude between grid cell centroids i and j and three indicator variables that take the value one if grid cells i and j are both intersected by a waterway, are both located on the Mediterranean Sea, and are both part of the same biome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

around 28%. When additionally accounting for geographical features in column (3), the point estimate remains stable. The geography controls include absolute difference in latitude between grid cell centroids, an indicator capturing whether both grids cells have access to a river or coast, an indicator for joint access to the Mediterranean Sea, and an indicator that takes the value one if grid-cell pairs share the same biome (i.e., the same biological community).

As illustrated in Section 3, Roman effective distance is correlated with geodesic distance. Ceteris paribus, the shipping costs within the Roman transport network increase when regions are further apart. To isolate the portion of the variation in Roman effective distance that is Roman-era specific, we therefore control for geodesic distance in column (4). While we find that—consistent with the trade literature—trade intensity rapidly declines with geodesic distance, the coefficient of Roman effective distance remains statistically significant and sizeable. This documents that variation in Roman effective distance is to a large part driven by Roman-era specific factors, i.e., the *combination* of the structure of the new transport network ($\mathbf{N}^{Rome}$) and the mode-specific freight rate differentials ($\boldsymbol{\alpha}^{Rome}$). The finding also accords well with historical narrative that indicates that the transport routes were highly non-linear in geographical distance (see Section 2). Traders made substantial detours to reach and make avail of more cost-effective

the Roman transport network, we cannot exploit time variation in our analysis. However, as mentioned previously, our results remain qualitatively unaltered if we run our regressions at the production-site level and include production site fixed effects. These dummies account, to a certain extent, for differences in timing, as production sites were operating at different times.

transport modes.

The results in columns (1)–(4) document that transport costs—measured by Roman effective distance—influenced the pattern of Roman trade. In columns (5)–(6) we show that our estimates do not conflate other aspects of connectivity within the Roman network. Accounting for the distance-wise shortest path or the time-wise shortest path changes the point estimate of Roman effective distance little. The coefficient of geodesic distance, on the other hand, becomes non-significant due to the high collinearity with the two measures.

The magnitude of the coefficients for the Roman effective distance elasticity in Table 2 is similar, although somewhat larger, compared to the pre-modern geographic-distance elasticity of trade of -1.9 reported in Barjamovic et al. (2019). Their estimates are based on commercial records produced by Assyrian merchants during the Bronze Age. Donaldson (2018) reports an effective-distance elasticity estimate of around -1.6 for 19th and 20th-century India. In a meta analysis of papers that estimate the effects of distance on trade for the current day, Disdier and Head (2008) find an average elasticity of -0.9, with 90% of the coefficients lying between -1.55 and -0.28.[31] Regardless of the controls included, our point estimates for the Roman era are at the upper end or above intervals estimated for later periods. This implies that the importance of distance has declined over time, which is in line with the common perception of decreasing transport costs and increased globalisation (see for example Bergstrand, Larch and Yotov, 2015).

The fact that a gravity-type relationship holds for Roman trade in terra sigillata implies that we observe regional specialisation in products or product varieties, which, in turn, leads to exchange of products or varieties, i.e., trade between regions. Many prominent theoretical underpinnings of the gravity model build on the existence of products or product varieties which induces intra-industry trade (see Anderson, 1979; Eaton and Kortum, 2002, for examples). Hence, such a framework fits well to the nature of our terra sigillata data, where product type and quality likely vary across production sites (see Section 2). [32]

Summing up, the results presented in Table 2 show that the creation of the Roman transport network and resulting differences in interregional costs of shipping goods strongly influenced the contemporary geography of trade.

5.2 From past to present

Historical narratives indicate that the Roman transport network continued to influence the trade patterns at least until the Industrial Revolution. Roman roads, for example, were maintained

[31] Note that many surveyed studies proxy trade costs by geographical distance, whereas our measure is effective distance.

[32] All theoretical foundations of the gravity model build on the assumption that there are many more goods than factors, which allows for complete specialization in different products or product varieties across countries (see Feenstra, 2016, p. 133). Gravity is consistent with perfect competition (see Eaton and Kortum, 2002; Eaton, Kortum and Sotelo, 2013) and monopolistic competition (see Anderson, 1979; Bergstrand, 1985) as well as a constant-elasticity of substitution utility function allowing for love-of-variety. However, the assumptions about trade-incentive-generating differences vary: Anderson (1979) and Bergstrand (1985) assume same productivities across countries, but allow for some monopoly power, Eaton and Kortum (2002) and Eaton, Kortum and Sotelo (2013) assume productivity differences across countries and perfect competition. But also a perfect competition Heckscher-Ohlin model with a continuum of goods may lead to a gravity-type relationship if factor prices differ (see Davis, 1995). In this case, countries specialise in different goods rather than varieties.

and continuously used during the Middle Ages (De Luca, 2016). Absent major innovations in transport technologies, this suggests that Roman-era specific differences in transport network connectivity persisted long after the collapse of the Roman Empire. In Appendix B we empirically support the historical evidence. Absent spatially disaggregated and temporally consistent information on economic interaction for the post-Roman period, we use two alternative proxies for market integration. The first proxy uses grid-cell-pair price correlations over the time period 1208–1790 and is constructed from data compiled in Federico, Schulze and Volckart (forthcoming). The second proxy is the time lag in the onset of the Black Death (1346–51) between grid cells and is constructed from information reported in Christakos et al. (2005). The use of this metric is motivated by the fact that the Plague spread along trade routes with merchants being the primary carriers of the disease (see, e.g., Cipolla, 1974; Biraben, 1975; Benedictow, 2006). Differences in the timing of onset can therefore be seen as measure of trade intensity during the Middle Ages (Boerner and Severgnini, 2014).

Using a regression setup analogous to Equation (1), we find that greater connectivity within the Roman transport network increases cross-regional price correlations and reduces time lags in the onset of the Black Death. This shows that differences in Roman effective distance continuously influenced the intensity of interregional economic integration from medieval times up to the beginning of the Industrial Revolution. Along with the sustained effect on economic interaction, greater connectivity within the Roman transport network arguably increased the flow of migrants and ideas and fostered cultural exchange. Such uninterrupted exchange reduces information asymmetries and thereby may influence interregional business link intensity until today (see, e.g., Guiso, Sapienza and Zingales, 2009; Leblang, 2010; Burchardi, Chaney and Hassan, 2019).

5.3 Roman transport network connectivity and economic integration today

To investigate whether the intensity of interregional business links today is influenced by differences in connectivity that emerged due to the creation of the Roman transport network, we continue to use regression equation (1), but employ the share of all subsidiaries located in grid cell j whose parent company is in grid cell i as outcome. Table 3 presents the results. In column (1), we estimate the effect of Roman effective distance conditional on geodesic distance, a same-country dummy, as well as the complete set of origin and destination fixed effects. When conditioning on geodesic distance, the Roman effective distance coefficient captures only the part of the transport network effect that is Roman-era specific. The point estimate of -0.475 illustrates that this Roman-era specific part of the transport cost variation strongly influences today's spatial firm ownership structure. Cross-regional firm link intensity decreases by 0.48% for each percent reduction in connectivity.[33] Column (1) also shows that geodesic distance exerts a strong negative effect on the intensity of economic interaction. Furthermore, the positive and statistically significant coefficient of the intra-national dummy unveils the existence of a home bias. That is, cross-regional firm ownership is more common within than across national

[33] Not conditioning on geodesic distance produces an Roman effective distance coefficient of -2.272.

borders.

Table 3: Roman transport network connectivity and interregional firm ownership today

Dependent Variable:	Share of Ownership Links (>25% Ownership)					
	Full Sample				Manufacturing	Service
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	-0.475***	-0.404***	-0.431***	-0.510***	-0.412***	-0.306**
	(0.075)	(0.077)	(0.088)	(0.090)	(0.090)	(0.123)
ln geodesic distance	-1.397***	-1.521***	-1.572***	-1.672***	-1.288***	-1.452***
	(0.054)	(0.060)	(0.085)	(0.078)	(0.070)	(0.083)
Intra-national ownership	1.623***	1.593***	1.596***	1.599***	1.386***	2.025***
	(0.102)	(0.107)	(0.107)	(0.108)	(0.107)	(0.168)
Joint duration under Roman rule (centuries)		0.479***	0.480***	0.485***	0.719***	0.586***
		(0.131)	(0.131)	(0.130)	(0.142)	(0.162)
ln network distance			0.075			
			(0.102)			
ln network time				0.249**		
				(0.104)		
Geography controls	No	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	731,823	731,823	731,823	731,823	602,597	470,736
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Dependent variable is the share of firms in cell j that are (partly) owned by firms located in cell i. 'Roman effective distance' represents the cost associated with shipping goods along the least cost path between grid cells, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'geodesic distance' represents the length in kilometres of the straight-line (as the crow flies) between grid cells. 'Intra-national ownership' is a dummy variable that captures whether grid cells i and j lie within the same country. Further controls are described in the notes of Table 2. The dependent variables in column 5 (Manufacturing) and column 6 (Service) include only ownership links whose parent firms belong in the specified sector. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Column (2) documents that the least-cost route coefficient remains stable when we augment the set of controls to include historical and geographical variables. As for Roman trade, the number of years jointly spent under Roman rule strongly increases cross-regional investment intensity. In column (3), we additionally control for network distance. This has little effect on the coefficient of Roman effective distance, documenting that our estimates do not capture differences in the distance-wise shortest route within the Roman network alone, but the combined effect of the network structure and Roman transport-mode-specific differences in transport costs. Accounting for differences in time-wise shortest path in column (4) also leaves the Roman effective distance coefficient relatively stable and highly statistically significant.

In the last two columns of Table 3, we analyse whether the effects of Roman effective distance varies between manufacturing and service parent firms. As the physical transport of goods is (relatively) unimportant for firms within the service industry, stark differences in the effect across sectors could provide an insight into whether the movement of goods today plays an important part in explaining our findings. There is, however, no indication that this is the case. Separately estimating regression Eq. (1) for manufacturing and service ownership companies produces similar, and statistically indistinguishable, point estimates. In Section 7, we discuss possible mechanisms that link Roman transport network connectivity to today's owner-

ship structure in more detail.

5.4 Robustness

A number of robustness checks show that our findings are not driven by specific assumptions or data construction choices. The results of all subsequently discussed exercises are presented in Appendix C in Table C.1 (Roman era) and C.2 (today). We first show that Roman effective distance reduces trade along the extensive as well as the intensive margin. We then document that our results remain stable if we include grid cells in our analysis that are not intersected by the Roman transport network or use the directed network (in which we differentiate between costs of up- and downstream river transportation) to calculate Roman effective distance. Alternative standard error clustering approaches also produce similar results. Additionally accounting for country-pair fixed effects in the current-day regressions changes little. We also show that our findings do not hinge upon the choice of a threshold in the ownership definition. Results are similar when we define ownership as having a minimum stake of 50% in another firm. Furthermore, we document that we obtain qualitative equivalent results if we use counts (i.e., number of pottery finds or number of ownership links) rather than shares as outcome variables (Tables C.5–C.6).

5.5 Extension

Before discussing the internal validity of our analysis, we reproduce our results using an alternative outcome. Recent evidence documents the importance of cross-border firm ownership—i.e., multinational firms—in explaining international business cycle transmissions (Cravino and Levchenko, 2017). Motivated by these findings, we investigate whether the Roman-era-specific effect on interregional firm ownership is also reflected in more synchronised business cycles. This auxiliary analysis, presented in Appendix D, may add to our understanding of the determinants of interregional contagion of economic shocks. Furthermore, using an alternative measure of economic integration (i.e., the intensity of business cycle transmission rather than business links) helps corroborate the findings presented above. Absent yearly grid-cell level data on GDP, we employ night-time light intensity as a proxy for regional income and compute the correlation in night-time lights growth between 1992 and 2013 for each grid-cell pair. As shown in Table D.2, income growth fluctuations between regions become less synchronised as connectivity within the Roman transport network decreases. Together with the results of Table 3, this illustrates that the Roman transport network continues to shape today's pattern of interregional economic integration in Western Europe. That is, today's intensity of economic interaction between two regions was (partly) determined by infrastructure investments and available transport technologies two millennia ago.

6 Threats to identification

As outlined in Section 3, variation in Roman effective distance is generated by two components: the structure of the network $\mathbf{N}^{Rome}$ (i.e., the Roman roads combined with the course of rivers and coastal routes) and the relative, mode-dependent, transport costs ($\boldsymbol{\alpha}^{Rome}$). The validity of our estimation strategy hinges on the assumption that—conditional on controls—the combination of these two components captures Roman-era-specific variation in transport costs and is uncorrelated with factors in the error term that influence outcome variables. There are two main threats to the validity of this assumption. The first is that the connectivity within the Roman network is influenced by pre-existing patterns of exchange and therefore is endogenous. The second is that Roman effective distance is spuriously capturing the effects of geography. Below, we apply a variety of complementary approaches to mitigate these concerns and document that our estimates are, in fact, capturing Roman-era-specific effects.

6.1 Pre-Roman interaction

We address the worry that connectivity within the Roman transport network is endogenous by testing if Roman effective distance predicts pre-existing patterns of interaction. To this end, we use a variety of measures for (socio)-economic exchange.

Goods trade The first set of measures for pre-Roman interaction is based on archaeological artefacts traded during the Neolithic and Bronze age. In analogy to the terra sigillata data, we define excavation sites as destinations and derive information on origins from provenance studies undertaken in the archaeological literature. For our empirical analysis, we draw on two existing databases of pre-Roman trade. These contain information on Neolithic axeheads that were primarily made from jade, for which the mining sites can be identified. Together the two sources include ca. 3,700 artefacts (Pétrequin et al., 2012; Schauer et al., 2020). We complement the existing databases with our own collection. From a variety of academic sources, we collect information on origin and destination of an additional 3,744 metal-based artefacts that were exchanged during the Bronze Age. These include weapons, tools, and jewellery made from bronze, copper, and silver (for further information, see Appendix E).

In analogy to our main analysis, we focus on Western Europe and aggregate the individual flows to the grid-cell-pair level by weighting all observations equally. That is, we interpret each find (irrespective of the type of good and data source) as one interaction. In Panel A of Table 4 we use regression model (1) and test if Roman effective distance influences pre-Roman interaction. Column (1) conditions on geodesic distance while geographic controls are added in column (2). In both specifications, the coefficient of geodesic distance is highly statistically significant. The estimates imply that trade decreases by more than 2% when geodesic distance increases by 1%. On the other hand, the point estimates of Roman effective distance—i.e., the Roman transport-technology-specific cost of shipping goods within the not yet existing Roman transport network—are small and non-significant. To investigate whether aggregation across goods and databases masks important heterogeneity, we run separate regressions for each of

the three databases in columns (3)–(5). Reassuringly, the picture remains consistent. Across all datasets, the coefficient of Roman effective distance is imprecisely estimated (and even positive in two instances), while geodesic distance continues to exert a strong negative effect on pre-Roman intensity of exchange. To address possible concerns related to low numbers of observations, we extend our dataset to include artefacts for which only the origin falls into Western Europe. To integrate these finds into the analysis, we assign their destinations to the nearest grid cell within Western Europe and weight these finds using the inverse distance (see Appendix E). This almost doubles the sample size. The pattern of results, however, does not change: Roman transport network connectivity does not predict intensity of interaction in pre-Roman times.

Cultural diffusion The second set of measures for pre-Roman interaction is based on burial and other cultural practices of the Neolithic and Bronze Age. Archaeologists generally agree that the spatio-temporal diffusion of these traditions took place by way of cultural exchange, including trade and migration (Cummings, Midgley and Scarre, 2015, p. 825 ff., Paulsson, 2019, Holst, 2013, p. 117, Childe, 1958, p. 123 ff., Childe, 1930, p.173 ff.). The occurrence of the same burial practice in two regions therefore implies economic and social interaction. Based on this insight, we use an indicator variable that takes the value of one if the same type of burial site, namely dolmen, chambered cairn, or round barrow, is found in grid i and grid j to reflect the extensive margin measure of interaction. Analogously, we construct a dummy that captures whether menhirs are located in both cells (see Appendix E for more information). These standing stones mark locations that are associated with a variety of cultural and religious practices and are an indicator of cultural and religious conformity across space (Walkowitz, 2003, p. 7). Due to the binary nature of the outcome variables, we estimate regression model (1) using OLS rather than PPML.[34]

Panel B of Table 4 presents the results. They show that while geodesic distance deterred cultural exchange during the Neolithic and the Bronze Age, Roman effective distance did not (columns 1–4). The same pattern of results prevails when we use a dummy for the (concurrent) presence of Celtic settlements during the La Tène culture in Gaul (Oppida) as proxy for interaction during the Iron Age (column 5). The use of this indicator is motivated by the observation that the Celtic culture spread across Europe via migration during the Iron Age. Taken together, the results presented in Table 4 provide strong evidence that connectivity within the Roman transport network does not reflect pre-existing patterns of interaction.

[34] That is, we estimate the following regression equation using OLS: $\mathcal{X}_{ij} = \delta \ln LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij} + T'_{ij}\gamma + \beta_i + \beta_j + \varepsilon_{ij}$. If we would use the number of burial sites found in each location and construct a variable thereof, we could also estimate the model in multiplicative form using PPML. However, we believe that the number of burial sites is not directly informative about the intensity of cultural exchange.

Due to the fact that grid cells without any indication of the respective cultural practice will result in an indicator that is zero for every grid-cell pair, the corresponding fixed effects perfectly explain these observations, which in the case of PPML leads to non-existence of the estimates as the first-order conditions corresponding to the fixed effects can never be satisfied (see Santos Silva and Tenreyro, 2010).

Table 4: Pre-Roman interaction

Panel a: Bilateral Trade in Goods (shares)

	All Goods	All Goods	Alpine Jade	British Jade	Metal Goods	All Goods
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	0.264	-0.252	-0.262	1.068	0.745	-0.136
	(0.361)	(0.376)	(0.488)	(0.920)	(1.256)	(0.335)
ln geodesic distance	-2.503***	-2.071***	-4.067***	-1.158***	-2.993***	-1.846***
	(0.186)	(0.155)	(0.794)	(0.177)	(1.036)	(0.152)
Baseline controls	No	Yes	Yes	Yes	Yes	Yes
Additional destinations assigned to grid	No	No	No	No	No	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,923	7,923	2,026	425	376	15,143
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Panel b: Diffusion of culture

	Neolithic			Bronze Age	Iron Age
	Both Cells Dolmen (Megalithic)	Both Cells Chambered Cairns (Megalithic)	Both Cells Menhir (Megalithic)	Both Cells Round Barrows (Tumulus)	Both Cells Oppida
	(1)	(2)	(3)	(4)	(5)
ln Roman effective distance	-0.009	0.002*	0.002	0.011	-0.002
	(0.008)	(0.001)	(0.010)	(0.007)	(0.003)
ln geodesic distance	-0.081***	-0.003***	-0.042***	-0.017***	-0.007***
	(0.006)	(0.001)	(0.005)	(0.004)	(0.002)
Mean dep. var.	0.135	0.002	0.158	0.030	0.014
Baseline controls	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes
Observations	407,253	407,253	407,253	407,253	407,253
Estimator	OLS	OLS	OLS	OLS	OLS

Notes: This table reports estimates of Equation (1) using the PPML estimator (panel a) and the OLS estimator (panel b). Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Dependent variables in Panel A are shares of finds in cell j that originates from cell i. In column (3) finds are Alpine jade (Neolithic) from Pétrequin et al. (2012); in column (4) finds are British axeheads (Neolithic) from Schauer et al. (2020); in column (5) finds are metal artefacts from our own data collection exercise; in columns (1)–(2) and (6) finds are all three types of goods combined. For more details see Appendix E. Dependent variables in Panel B are indicator variables taking the value one if a given feature is observed in grid cell i and j. The Neolithic features are: dolmen (column 1), chambered cairns (column 2), and menhirs (column 3). The Bronze age feature is: round barrow (column 4). The Iron age feature is: oppidum. For more details see Appendix E. 'Roman effective distance' represents the cost associated with shipping goods along the least cost path between grid cells, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'geodesic distance' represents the length in kilometres of the straight-line (as the crow flies) between grid cells. Baseline controls correspond to column 2 in Table 2 and are described in the respective table notes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.2 Instrumental variable approach

Notwithstanding the evidence just presented, there may be remaining concerns that connectivity in the Roman transport network is endogenous. One particular worry is that the construction of Roman roads—the Roman-specific component in the structure of the multi-modal transport network—was influenced by unobserved (socio-)economic factors. To address this, we develop an instrumental variable strategy. The approach—outlined in more detail in Appendix F—is based on the historical evidence that construction and design of Roman roads was primarily

determined by military-strategic aims. Roads were primarily built to facilitate supply shipments and bringing in reinforcements from Rome into newly annexed regions. To minimise construction cost and travel times for troops, roads often followed straight lines over long distances (cf. Section 2). Building on these insights, we construct a road network that connects the capital Rome to the locations of Roman battlefields using the Gabriel graph (Gabriel and Sokal, 1969). This hands-off approach ensures that road connections are not drawn based on subjective (or arbitrary) criteria. Intuitively, the Gabriel graph algorithm produces a road network in which neighbouring locations are connected using straight-line segments. The direct connection between strategic-military nodes implies that neither economic conditions nor geographic characteristics influence the path of these roads.

To construct our instrument, we replace the Roman roads in the multi-modal network with the Gabriel roads and identify least cost paths and associated costs in analogy to the procedure described in Section 3. Transshipment between road and the two other modes of transport is allowed at intersections. Because the Gabriel road network is less dense than the actual Roman road network, fewer grid cells are intersected by the IV network. For non-intersected grid cells we cannot predict Roman effective distance using the IV. Compared to the main analysis, the number of observations is thus reduced. The exclusion restriction would be violated if the location of battles is 'chosen' such that the resulting connectivity within the IV network is endogenous to geographical connectivity or pre-existing patterns of exchange. We believe that this is unlikely to be the case because the routing of roads as straight lines in the Gabriel graph ignores any economic or geographic characteristics. Furthermore, the concern that battles are potentially more likely to occur in cells that are valuable trading partners is addressed by the inclusion of origin and destination fixed effects.

Table 5 presents results of the instrumental variable approach for the Roman era (columns 1–3) and today (columns 4–6). For reference, column (1) replicates the results of our preferred specification (column 4 of Table 2). In column (2), we run the same specification using the sample restricted to grid-cell pairs that are intersected by the IV network. The results are very similar. Crucially, column (3) shows that the IV procedure also produces a negative and statistically highly significant point estimate of Roman effective distance on Roman trade. In columns (4)–(6) we repeat the above exercise using current-day ownership link intensity as dependent variable (replicating the preferred specification from column (2) in Table 3). Again, the IV estimates in column (6) confirm the relationship of Roman effective distance with business links. Compared to the standard PPML estimates in columns (2) and (5), the IV coefficients are somewhat larger. One potential explanation is that the PPML estimates suffer from attenuation bias. Classical measurement error may arise from imprecise maps of Roman roads or differences in the quality of roads, biasing our estimates towards zero. Combined, the results of Table 5 show that military-strategic objectives were a key determinant of Roman road construction and that our findings are not driven by endogenous or geography-driven routing of Roman roads.

Table 5: Instrumental variable approach

Dependent Variable:	Share of Pottery Finds			Share of Ownership Links (>25% Ownership)		
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	-1.493***	-1.789***	-2.651***	-0.404***	-0.322***	-0.645***
	(0.542)	(0.457)	(0.551)	(0.077)	(0.090)	(0.138)
ln geodesic distance	-0.655***	-0.866***	-0.508	-1.521***	-1.540***	-1.404***
	(0.230)	(0.314)	(0.328)	(0.060)	(0.067)	(0.073)
Same country	No	No	No	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	IV	IV	Full	IV	IV
Observations	22,839	15,698	15,698	731,823	442,183	442,183
Estimator	PPML	PPML	IV PPML	PPML	PPML	IV PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator (columns 1–2 and 4–5) and the IV PPML estimator (columns 3 and 6). Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Dependent variables are the share of terra sigillata finds in cell j that originates from cell i (columns 1–3) or the share of firms in cell j that are (partly) owned by firms located in cell i (columns 4–6). Column (1) shows the baseline specification from column (4) in Table 2; column (2) shows results from this specification in the sample for which the instrumental variable is available; column (3) shows results where effective distance is instrumented with a measure of effective distance that replaces roads with straight-line segments from a Gabriel graph, as described in detail in Appendix F. Columns (4–6) repeat the procedure of columns (1–3) based on the specification in column (2) in Table 3. 'Roman effective distance' represents the cost associated with shipping goods along the least cost path between grid cells, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'geodesic distance' represents the length in kilometres of the straight-line (as the crow flies) between grid cells. Baseline controls correspond to column 2 in Table 2 and are described in the respective table notes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.3 Controlling for geographic connectivity

The routing of roads is only one component that generates variation in Roman effective distance. The IV exercise therefore does not (entirely) quash concerns related to the possibility that connectivity within the network is correlated with underlying geography (access to rivers and coastlines in particular). To document that this is not the case, we control for a variety of geography-based least-cost path measures. These measures, described in Appendix G, are designed to capture general, Roman-infrastructure-unrelated, costs of transporting goods and people between regions during the pre-industrial era. Importantly, all least-cost path measures allow for transport over land, river, and sea. In both the Roman and the current-day analysis, coefficients of Roman effective distance remain stable irrespective of whether we model the costs for geography-based least-cost paths in terms of time or energy expenditure (see Table G.1–G.2). These tables also show that coefficients change very little when we augment the regressions with a wide range of additional geographical and climatic aspects.[35]

6.4 Falsification

The exercises presented above suggest that our estimates specifically capture the effects of Roman transport network connectivity rather than pre-existing patterns of exchange or geographic proximity. As a final approach to underpin the credibility of our results, we conduct a falsifica-

[35]This set of additional controls encompasses the absolute difference in longitude, elevation, ruggedness, agricultural suitability, precipitation, temperature, and access to waterways. Measures of ruggedness along the straight line, indicators whether the straight line crosses a river or coastline, an indicator for location on the same watershed, and three indicators that capture whether the least cost path runs via any river, road and sea segments, respectively.

tion test. Due to data availability, the test is restricted to business link intensity as outcome.

For the purpose of this exercise and contrary to all the historical and empirical evidence presented above, we assume the following:

(i) The routing of roads is determined by geographical features, meaning that the structure of the Roman transport network is simply reflecting underlying geography. The layout of the network would therefore not be Roman specific.

(ii) The vector of relative transport costs $\boldsymbol{\alpha}^{Rome}$ represents general historical transport cost ratios and would therefore not be Roman specific.

If (i) *and* (ii) hold, Roman effective distance captures variation in connectivity that is not specific to the Roman era but rather determined by geographical features and universally applicable historical transport technologies. This implies that Roman effective distance should also predict the intensity of interregional interaction outside of the former Roman Empire.

Unfortunately, we lack (detailed) information on the structure of historical road networks for non-Roman areas. However, assuming that today's primary roads (highways) largely follow historical routes, we can use the structure of the current road network as a proxy for the historical one. We regard this assumption as plausible given the quantitative and qualitative evidence that current highways follow historical paths, both in regions within and beyond the Roman Empire (e.g. Garcia-López, Holl and Viladecans-Marsal, 2015; Percoco, 2015; Redding and Turner, 2015; Hitchner, 2012). We therefore combine today's highway network (as a proxy for the historical road network) with rivers and coastlines into a multi-modal transport network that spans regions of Europe that were part of the Roman Empire and those that were not.[36] Based on the historical cost ratios $\boldsymbol{\alpha}^{Rome}$, we identify the least cost paths between grid-cell pairs and compute the corresponding Roman effective distance in analogy to the procedure described in Section 3. This produces a measure of bilateral connectivity that is constructed in an identical way for regions within and outside Western Europe. This measure allows us to test whether the relationship between Roman effective distance and business link intensity differs between European areas located inside and outside of Rome applying regression equation (1) (see Appendix H for more details). For regions once integrated into Rome, we unambiguously expect that lower transport-network connectivity is associated with lower business link intensity. For regions beyond Roman influence, we expect that higher effective distance deters business link formation only if this measure captures variation in connectivity that is not specific to the Roman-era. It is important to note that the results of the falsification test are not driven by the fact that one area was occupied by the Romans while the other was not. We run separate regressions for the regions inside and outside the empire. Furthermore, we are not looking at levels of interaction, but at differences in the intensity of bilateral interaction as transport costs within the (hypothetical) network vary. Table 6 reports the results of the falsification exercise. In column (1), we find that business link intensity between grid-cell pairs located within the border of Rome decreases with Roman effective distance. Compared to our main results (Table 2, column 2), the coefficient size is marginally smaller and less precisely estimated. This is unsurprising given that we use

[36] Data on primary roads come from ESRI (2020), information on course of large rivers is taken from WISE (2020).

Table 6: Falsification exercise

Dependent Variable:	Share of Ownership Links (>25% Ownership)	
	Europe once part of Roman Empire	Europe never part of Roman Empire
	(1)	(2)
ln Roman effective distance	-0.327***	-0.035
	(0.086)	(0.136)
ln geodesic distance	-1.559***	-1.361***
	(0.058)	(0.121)
Same country	Yes	Yes
Geography controls	Yes	Yes
Destination FEs	Yes	Yes
Origin FEs	Yes	Yes
Observations	723,323	414,451
Estimator	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Column (1) uses a sample of grid cells located in Western Europe, i.e. the geographical scope of our analysis. Column (2) uses a sample of grid cells located in Europe outside of the former borders of Rome (for details, see Appendix H). Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Dependent variable is the share of firms in cell j that are (partly) owned by firms located in cell i, restricted to the respective regions of each sample. 'Roman effective distance' represents the cost associated with shipping goods along the least cost path between grid cells, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'geodesic distance' represents the length in kilometres of the straight-line (as the crow flies) between grid cells. Control variables are described in the notes of Tables 2 and 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the modern road network as proxy for the historical one. For areas outside of Rome, variation in Roman effective distance does not predict business link strength. The coefficient is close to zero and not statistically significant. These findings provide additional evidence that our estimates specifically reflect the effects of Roman transport network connectivity.

7 Channels connecting Roman connectivity and current integration

In the final part of the study, we analyse potential mechanisms underlying our findings. Guided by the recent literature, we focus on four potential channels: (i) persistence in interregional transport infrastructure connectivity, as well as Roman-network-induced similarity in (ii) production structures, (iii) preferences, and (iv) cultural values. Data on preferences are not available for Belgium and Luxembourg. The sample size therefore decreases from 731,823 to 674,805 observations. All subsequently used data is described in Appendix I.

7.1 Potential channels

The cost of transporting goods and people influences investment decisions of firms. If Roman transport network connectivity influences the (relative) accessibility of regions within today's transport networks, this could provide one explanation for our findings. To test for the plausibility of persistent transport network connectivity as mechanism we use two distinct measures of transport costs. The first is driving distance along the time-minimising route between grid cell centroids (extracted from Google Maps), which we interpret as capturing the cost of transport-

ing goods and people using today's road network.[37] This metric captures variation arising from distance in the road network and differences in the speed of transport associated with different technologies (i.e. motorways, rural roads, etc.).[38] The second measure specifically captures passenger-transport network connectivity. The focus on passenger transport links is motivated by recent studies showing that travel times strongly influence the intensity of cross-regional business connections (Giroud, 2013; Campante and Yanagizawa-Drott, 2018). Minimum travel time—our measure for passenger transport connectivity—between grid-cell-centre pairs is extracted from `rome2rio.com`. Within this multi-modal network, passengers are allowed to use any combination of public transport (bus, train, aeroplane).[39]

In Table 7, we investigate the plausibility of these two channels by testing if Roman effective distance influences accessibility within today's transport networks. To facilitate comparison, the dependent variables are standardised with mean zero and a standard deviation of one.[40] Column (1) shows that lower transport costs during Roman times are reflected in better accessibility within the road network today. However, the coefficient is small and estimated relatively imprecisely. A potential explanation for this finding is that the road network today is extremely dense and only allows for one mode of transportation. This implies that road network distances are highly correlated with geodesic distance. The amount of residual variation left to explain is therefore very small. For the multi-modal passenger transport, on the other hand, we find a large and statistically highly significant effect (column 2). This suggests that regions with historically stronger ties were connected more directly when new transport technologies became available (e.g., railways, aeroplanes, and highways). Thus, even though past and present multi-modal transport networks structurally differ in their layout and transport technologies, Roman-era-specific connectivity still explains patterns of bilateral accessibility today.

As a second potential channel, we investigate whether regions better connected within the Roman transport network have similar production structures. Continued economic interaction could, for example, have resulted in assimilation of industry structures and thereby facilitate cross-regional firm ownership (see Burchardi, Chaney and Hassan, 2019). Column (3) indeed indicates that production structures between regions become more dissimilar—as measured by an industry dissimilarity index based on Jaffe (1986)—when bilateral connectivity decreases (i.e., when Roman effective distance increases). Along with stimulating interregional trade, greater connectivity within the Roman transport network is likely to have affected the flow of migrants, ideas, and culture. This could have led to co-evolution and assimilation of preferences, values, and attitudes over the long run. Greater similarity in these fundamental determinants of economic interaction, in turn, can facilitate investment (Guiso, Sapienza and Zingales, 2009; Leblang, 2010; Burchardi, Chaney and Hassan, 2019). Firms, for example, may derive a competitive advantage from catering to multiple markets that exhibit a similar demand structure. Furthermore,

[37] Today, road transport is the dominant mode of shipping within Europe, accounting for 76% of the total volume of goods transported in 2017 (Eurostat, `http://bit.do/ModalSplit`).

[38] The two aspects—distance and time—are two main determinants of the overall road transport costs today (Persyn, Díaz-Lanchas and Barbero, 2020).

[39] We also allow for taxi rides when public transport is not available.

[40] Note that we use OLS instead of PPML because dependent variables can take negative values.

Table 7: Channels connecting Roman transport network connectivity and current integration

	ln Google driving distance (SD)	ln Rio2Rome travel time (SD)	Industry dissimilarity (SD)	First principal component preferences (SD)	First principal component attitudes (SD)	ln SCI (SD)
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	0.012*	0.405***	0.116***	0.112***	0.170***	-0.274***
	(0.007)	(0.025)	(0.026)	(0.021)	(0.024)	(0.025)
ln geodesic distance	1.389***	0.798***	0.041***	0.225***	0.168***	-0.588***
	(0.005)	(0.014)	(0.012)	(0.015)	(0.013)	(0.013)
Raw mean of dep. var.	6.962	6.241	0.800	0	0	6.518
SD of raw dep. var.	0.680	0.331	0.161	1.228	1.344	1.713
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	674,805	674,805	674,805	674,805	674,805	674,805
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Notes: This table reports estimates of Equation (1) using the OLS estimator. Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Each column uses a different dependent variable that serves a mechanism, such as current-day transport connectivity (columns 1–2), industry dissimilarity (column 3), cultural dissimilarity (column 4–5), and social connectedness (column 6). For details on the dependent variables, see the main text and Appendix I. Baseline controls correspond to column 2 in Table 3 and are described in the notes of Tables 2 and 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

similarity in preferences and values can reduce information frictions and coordination costs. To investigate whether differential connectivity within the Roman transport network explains variation in preferences and values across space, we draw on geocoded individual-level data from the Global Preferences Survey (GPS, Falk et al., 2018) and the European Values Study (EVS, EVS, 2016). The GPS contains information on six preferences: time preferences, risk preferences, positive and negative reciprocity, altruism, and trust. The EVS elicits information on human values and attitudes that are grouped into six categories: life, work, religion, family, politics and society, and nationalism. To bring the individual-level information to the grid cell level, we first purge the data of country fixed effects and then compute grid-cell-level means by averaging across all respondents that reside within a given cell.[41] We generate measures of cultural similarity across grid-cell pairs by computing the absolute difference in preferences and values between any two grid cells. For ease of exposition, we subsequently focus on the first principal components rather than the specific preferences and values. Column (4) shows that preferences become more dissimilar when Roman effective distance increases. Likewise, regions ill-connected within the Roman network exhibit greater disparities in attitudes and values (column 5). In Appendix I, we analyse the effects of Roman network connectivity separately for the individual preferences and value groups. We find that dissimilarity in the majority of aspects increases with Roman effective distance. This highlights the importance of considering socio-economic forces, in our case the transport-network-induced cumulative history of exchange between regions, when trying to understand why preferences and attitudes vary across regions.

Columns (1)–(5) shed light on the plausibility of specific mechanisms proposed by the recent literature. Clearly, there are further mechanisms that could underlie our reduced-form results. These include, for example, reduced genetic distance due to (network-connectivity-

[41] Note that the most detailed geographical information available on residence of respondents in both surveys is the NUTS 2 level. A detailed description of the data construction process, including the matching of respondents to grid cells, is provided in Appendix I.

induced) historical migration or simply increased familiarity and trust due to cumulative history of exchange. Such channels, while plausibly important, are inherently hard to measure and can therefore not be included in our analysis. However, one measure that potentially subsumes many potential mechanisms (including the ones discussed in columns (1)–(5) of Table 7) is the Social Connectedness Index (SCI), developed and described in detail in Bailey et al. (2018*a*). The SCI captures the link strength between two regions within the Facebook network and has been shown to influence investment flows (Bailey et al., 2018*b*). In the context of our analysis, social connectedness can (loosely) be interpreted as a composite index of revealed similarity. Assuming that differences in transport network infrastructure, preferences and values are likely to be reflected in the intensity of social ties, we expect that Roman transport network connectivity predicts differences in SCI. Column (6) shows that this is, in fact, the case. Social connections are less intense between region pairs that were ill-connected within the Roman network.

7.2 Relative importance of channels

Table 8: Accounting for potential channels

Dependent Variable:	Number of Ownership Links (>25% Ownership)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln Roman effective distance	-0.370***	-0.362***	-0.237***	-0.236***	-0.362***	-0.326***	-0.318***	-0.304***	-0.188***	-0.065
	(0.079)	(0.079)	(0.073)	(0.073)	(0.079)	(0.078)	(0.078)	(0.078)	(0.072)	(0.063)
ln Driving distance (SD)		-0.218*		-0.028					0.041	
		(0.127)		(0.130)					(0.124)	
ln Rome2Rio (SD)			-0.249***	-0.248***					-0.221***	
			(0.021)	(0.022)					(0.021)	
Industry dissimilarity (SD)					-0.122***				-0.117***	
					(0.024)				(0.023)	
Distance preferences (SD)						-0.217***		-0.138***	-0.113***	
						(0.021)		(0.021)	(0.021)	
Distance values (SD)							-0.246***	-0.179***	-0.170***	
							(0.028)	(0.029)	(0.027)	
ln SCI (SD)										1.140***
										(0.043)
ln geodesic distance	1.541***	-1.243***	-1.160***	-1.122***	-1.534***	-1.382***	-1.386***	-1.325***	-1.063***	-0.344***
	(0.062)	(0.168)	(0.072)	(0.158)	(0.062)	(0.066)	(0.063)	(0.066)	(0.152)	(0.063)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	674,805	674,805	674,805	674,805	674,805	674,805	674,805	674,805	674,805	674,805
Estimator	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Dependent variable is the share of firms in cell j that are (partly) owned by firms located in cell i. Each column adds explanatory variables that serve as mechanisms to explain the results in Table 3. Lower number of observations due to missing data for BEL and LUX. For details on these explanatory variables, see the main text and Appendix I. Baseline controls correspond to column 2 in Table 3 and are described in the notes of Tables 2 and 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the final step of our analysis, we assess the relative importance of the channels introduced above. In a horse race specification we regress the number of business links on Roman effective distance while adding the proxies for the various potential mechanisms. The results are reported in Table 8. In column (1), we run our preferred regression specification (see column (2), Table 3) on the restricted sample. This produces a point estimate of -0.370. The size of the Roman effective distance coefficient drops by 2% (column 2) and 36% (column 3) when

we account for differences in modern road transport costs and passenger transport accessibility. Combined, the two variables absorb 36% of the Roman transport network coefficient (column 4). This implies that continued interregional transport infrastructure connectivity—particularly bilateral passenger accessibility—is one reason why the Roman transport network influences today's spatial firm ownership structure. However, a substantial part of the main effect remains unexplained by this mechanism. As shown in column (5), differences in production structures do not help explain this gap. Compared to column (1), the Roman effective distance coefficient is reduced only marginally when we control for industry similarity.

Next, we analyse the importance of preference and value similarity as mediating channels.[42] Including preference disparities into the regression setup reduces the Roman effective distance coefficient by 12% (column 6), while it drops by 14% when differences in attitudes and values are accounted for in column (7). Combined, differences in preferences and values account for 18% of the Roman effective distance coefficient (column 8). Our earlier findings suggest that the Roman transport network created a new pattern of bilateral interregional (socio-)economic interaction which, over time, led to an increase in preference and value similarity. This, in turn, can (partly) explain variation in cross-regional investment intensity. In column (9), we simultaneously add all potential mechanisms. Together, they absorb half of the Roman transport network effect on today's spatial firm ownership structure.

In the last column of Table 8, we account for the SCI rather than specific mechanisms. This index absorbs a large part of the variation in Roman transport network connectivity. The point estimate of Roman effective distance drops by 82% and is no longer statistically significant at conventional confidence levels. This result supports our argument that by creating and intensifying socio-economic exchange, the Roman transport network influences business link intensity today. Due to the composite nature of the SCI, however, we cannot derive any additional insights regarding specific mechanisms underlying our reduced-form effects.

8 Conclusion

This paper aimed at analysing the effects of the Roman transport network on economic integration in the past and the present. We document that the creation of the network generated a new pattern of interregional trade within Western Europe that persisted long after the fall of the Roman Empire. Along with continued economic integration, greater connectivity also led to convergence in values and attitudes. This network-induced assimilation in fundamental determinants of economic interaction, in turn, helps to explain patterns of economic interaction today. Similarly, despite the fundamental changes in available transport technologies, today's transport network connectivity patterns reflect ancient connectivity patterns. Partly as a result of these effects, business links are much stronger between regions that were better connected within the Roman network, illustrating the long-lasting and multifaceted consequences of infrastructure investments. Current barriers to integration are thus an outcome of historical integration. There-

[42] In Table I.3 of Appendix I, we add all individual measures of preferences and values and attitudes instead of their principal components. The results are very similar.

fore, policy makers need to be aware of, and take into account, the long-run consequences of public infrastructure investments. These investments can create or reshape networks in which the transmission of positive and negative shocks is more pronounced.

References

Adams, Collin. 2012. "Transport." In *The Cambridge Companion to the Roman Economy*, edited by Walter Scheidel, 218–240. Cambridge University Press.

Ahern, Kenneth R., Daniele Daminelli, and Cesare Fracassi. 2015. "Lost in Translation? The Effect of Cultural Values on Mergers around the World." *Journal of Financial Economics* 117 (1): 165–189.

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

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018*a*. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives* 32 (3): 259–80.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018*b*. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy* 126 (6): 2224–2276.

Bakker, Jan David, Stephan Maurer, Jörn-Steffen Pischke, and Ferdinand Rauch. forthcoming. "Of Mice and Merchants: Trade and Growth in the Iron Age." *Review of Economics and Statistics* , (24825).

Baldwin, Richard E., and James Harrigan. 2011. "Zeros, Quality, and Space: Trade Theory and Trade Evidence." *American Economic Journal: Microeconomics* 3 (2): 60–88.

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

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. 2017. "Roads, Railroads, and Decentralization of Chinese Cities." *Review of Economics and Statistics* 99 (3): 435–448.

Benedictow, Ole Jørgen. 2006. *The Black Death, 1346-1353: The Complete History.* Boydell Press.

Berechman, Joseph. 2003. "Transportation-economic Aspects of Roman Highway Development: The Case of Via Appia." *Transportation Research Part A: Policy and Practice* 37 (5): 453–478.

Bergstrand, Jeffrey H. 1985. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence." *Review of Economics and Statistics* 67 (3): 474–481.

Bergstrand, Jeffrey H., Mario Larch, and Yoto V. Yotov. 2015. "Economic Integration Agreements, Border Effects, and Distance Elasticities in the Gravity Equation." *European Economic Review* 78: 307–327.

Biraben, Jean-Noël. 1975. *Les Hommes et la Peste en France et dans les Pays Europeens et Mediteraneens.* Mouton.

Boerner, Lars, and Battista Severgnini. 2014. "Epidemic Trade." London School of Economics Economic History Working Papers 212.

Bowman, Alan, and Andrew Wilson. 2009. "Quantifying the Roman Economy: Integration, Growth, Decline?" In *Quantifying the Roman Economy: Methods and Problems*, edited by Alan Bowman and Andrew I. Wilson, 3–86. Oxford: Oxford University Press.

Burchardi, Konrad B., Thomas Chaney, and Tarek A. Hassan. 2019. "Migrants, Ancestors, and Foreign Investments." *Review of Economic Studies* 86 (4): 1448–1486.

Burstein, Ariel, Christopher Kurz, and Linda Tesar. 2008. "Trade, Production Sharing, and the International Transmission of Business Cycles." *Journal of Monetary Economics* 55 (4): 775–795.

Campante, Filipe, and David Yanagizawa-Drott. 2018. "Long-Range Growth: Economic Development in the Global Network of Air Links." *The Quarterly Journal of Economics* 133 (3): 1395–1458.

Campbell, Brian. 2012. *Rivers and the Power of Ancient Rome.* UNC Press Books.

Carreras, Cesar, and Pau De Soto. 2013. "The Roman Transport Network: A Precedent for the Integration of the European Mobility." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 46 (3): 117–133.

Chaney, Thomas. 2014. "The Network Structure of International Trade." *American Economic Review* 104 (11): 3600–34.

Chen, M. Keith. 2013. "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." *American Economic Review* 103 (2): 690–731.

Chevallier, Raymond. 1972. *Les Voies Romaines.* Paris: Armond Colin.

Childe, Gordon. 1930. *The Bronze Age.* Cambridge University Press.

Childe, Gordon. 1958. *The Prehistory of European Society.* Penguin Books.

Christakos, George, Ricardo A. Olea, Marc L. Serre, Lin-Lin Wang, and Hwa-Lung Yu. 2005. *Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death.* Springer.

Cipolla, Carlo M. 1974. "The Plague and the Pre-Malthus Malthusians." *Journal of European Economic History* 3 (2): 277–284.

Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer. 2005. "The Trade-Creating Effects of Business and Social Networks: Evidence from France." *Journal of International Economics* 66 (1): 1–29.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson Estimation with High-Dimensional Fixed Effects." *The Stata Journal* 20 (1): 95–115.

Cravino, Javier, and Andrei A. Levchenko. 2017. "Multinational Firms and International Business Cycle Transmission." *The Quarterly Journal of Economics* 132 (2): 921–962.

Cummings, Vicki, Magdalena S. Midgley, and Chris Scarre. 2015. "The Oxford Handbook of Neolithic Europe." edited by Chris Fowler, Jan Harding and Daniela Hofmann, Chapter Chambered Tombs and Passage Graves of Western and Northern Europe. Oxford University Press.

Dalgaard, Carl-Johan, Nicolai Kaarsen, Ola Olsson, and Pablo Selaya. 2018. "Roman Roads to

Prosperity: Persistence and Non-Persistence of Public Goods Provision." mimeo.

Darling, Margaret J. 1998. "Samian from the City of Lincoln: A Question of Status?" In *Form and Fabric. Studies in Rome's material past in Honour of B.R. Hartley*, edited by J. Bird, 169–178. Oxbow Monograph.

Davies, Hugh E. H. 1998. "Designing Roman Roads." *Britannia* 29: 1–16.

Davis, Donald R. 1995. "Intra-Industry Trade: A Heckscher-Ohlin-Ricardo Approach." *Journal of International Economics* 39: 201–226.

De Benedictis, Luca, Vania M. Licio, and Anna M. Pinna. 2018. "The Long-term Effects of the Historical Roman Road Network: Trade Costs of Italian Provinces." Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia Working Paper CRENoS 201801.

De Luca, Giuseppe. 2016. "Infrastructure Financing in Medieval Europe." In *Infrastructure Finance in Europe: Insights into the History of Water, Transport, and Telecommunications*, edited by Youssef Cassis, Giuseppe De Luca and Massimo Florio, Chapter 1. OUP Oxford.

de Souza, Philip. 2002. *Piracy in the Graeco-Roman World.* Cambridge University Press.

Dijkstra, Edsger W. 1959. "A Note on Two Problems in Connexion with Graphs." *Numerische Mathematik* 1 (1): 269–271.

Disdier, Anne-Célia, and Keith Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *Review of Economics and Statistics* 90 (1): 37–48.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review* 108 (4–5): 899–934.

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

Duncan-Jones, Richard. 1974. *The Economy of the Roman Empire. Quantitative Studies.* Cambridge University Press.

Duranton, Gilles, and Matthew A. Turner. 2012. "Urban Growth and Transportation." *Review of Economic Studies* 79 (4): 1407–1440.

Duranton, Gilles, Peter M. Morrow, and Matthew A. Turner. 2014. "Roads and Trade: Evidence from the US." *Review of Economic Studies* 81 (2): 681–724.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741–1779.

Eaton, Jonathan, Samuel Kortum, Brent Neiman, and John Romalis. 2016. "Trade and the Global Recession." *American Economic Review* 106 (11): 3401–3438.

Eaton, Jonathan, Samuel S. Kortum, and Sebastian Sotelo. 2013. "International Trade: Linking Micro and Macro." In *Advances in Economics and Econometrics: Tenth World Congress*, Vol. 3 of *Econometric Society Monographs*, edited by Daron Acemoglu, Manuel Arellano and Eddie Dekel. Cambridge University Press.

Erdkamp, Paul. 2013. "The Food Supply of the Capital." In *The Cambridge Companion to Ancient Rome*, edited by Paul Erdkamp, 262–264. Cambridge University Press.

ESRI. 2020. "World Roads." http://bit.do/worldroads.

EVS. 2016. "European Values Study 2008." 4th wave, Integrated Dataset. GESIS Data Archive, Cologne, Germany, ZA4800 Data File Version 4.0.0.

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

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–1692.

Federico, Giovanni, Max-Stephen Schulze, and Oliver Volckart. forthcoming. "European Goods Market Integration in the Long Run." *Journal of Economic History*.

Feenstra, Robert C. 2016. *Advanced International Trade: Theory and Evidence.* 2 edition. Princeton, New Jersey: Princeton University Press.

Feyrer, James. 2019. "Trade and Income–Exploiting Time Series in Geography." *American Economic Journal: Applied Economics* 11 (4): 1–35.

Finley, Moses I. 1999. *The Ancient Economy.* Updated edition by I. Morris edition. University of California Press.

Fitzpatrick, Andrew. 1985. "The Distribution of Dressel 1 Amphorae in Northwest Europe." *Oxford Journal of Archaeology* 4 (3): 305–340.

Flückiger, Matthias, and Markus Ludwig. 2019. "Transport Infrastructure, Growth and Persistence: The Rise and Demise of the Sui Canal." *Canadian Journal of Economics* 52 (2).

Franconi, Tyler. 2014. "The Economic Development of the Rhine River Basin in the Roman Period (30 BC–AD 406)." PhD diss. Oxford University, UK.

Fulford, Michael. 2018. "Procurators' Business? Gallo-Roman Sigillata in Britain in the Second and Third Centuries AD." 301–326.

Furger, Alex, and Sabine Deschler-Erb. 1992. *Das Fundmaterial aus der Schichtenfolge beim Augster Theater: Typologische und osteologische Untersuchungen zur Grabung Theater-Nordwestecke 1986/87.* Römermuseum Augst.

Gabriel, K. Ruben, and Robert R. Sokal. 1969. "A New Statistical Approach to Geographic Variation Analysis." *Systematic Biology* 18 (3): 259–278.

Galor, Oded, and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review* 106 (10): 3064–3103.

Garcia-López, Miquel-Àngel, Adelheid Holl, and Elisabet Viladecans-Marsal. 2015. "Suburbanization and Highways in Spain When the Romans and the Bourbons Still Shape Its Cities." *Journal of Urban Economics* 85: 52–67.

Garmendia, Aitor, Carlos Llano, Asier Minondo, and Francisco Requena. 2012. "Networks and the Disappearance of the Intranational Home Bias." *Economics Letters* 116 (2): 178–182.

Giroud, Xavier. 2013. "Proximity and Investment: Evidence from Plant-Level Data." *The Quarterly Journal of Economics* 128 (2): 861–915.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Ex-

change?" *The Quarterly Journal of Economics* 124 (3): 1095–1131.

Hartley, Brian R., Brenda M. Dickinson, Geoffrey B. Dannell, Michael Fulford, Allard W. Mees, Paul Tyers, and Rosemary Wilkinson. 2008. *Names on Terra Sigillata. an Index of Makers' Stamps & Signatures on Gallo-roman Terra Sigillata (Samian Ware).* Vol. 1-9. University of London, Institute of Classical Studies.

Hitchner, R. Bruce. 2003. "Roman Empire." In *The Oxford Encyclopedia of Economic History*, Vol. 4, edited by Joel Mokyr, 397–401. Oxford University Press.

Hitchner, R. Bruce. 2012. "Roads, Integration, Connectivity, and Economic Performance in the Roman Empire." In *Highways, Byways, and Road Systems in the Pre-Modern World*, edited by Richard J. A. Talbert Susan E. Alcock, John Bodel, 222–234. Wiley Blackwell.

Holst, Mads Kähler. 2013. "The Oxford Handbook of the European Bronze Age." edited by Harry Fokkens and Anthony Harding, Chapter Burials. Oxford University Press.

Hopkins, Keith. 1980. "Taxes and Trade in the Roman Empire (200 Bc–ad 400)." *The Journal of Roman Studies* 70: 101–125.

Horden, Peregrine, and Nicholas Purcell. 2000. *The Corrupting Sea: A Study of Mediterranean History.* Oxford: Blackwell Publishers.

Hornung, Erik. 2015. "Railroads and Growth in Prussia." *Journal of the European Economic Association* 13 (4): 699–736.

Jaffe, Adam. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review* 76 (5): 984–1001.

Jedwab, Remi, and Alexander Moradi. 2016. "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa." *Review of Economics and Statistics* 98 (2): 268–284.

Johnson, Noel D., and Mark Koyama. 2017. "Jewish Communities and City Growth in Preindustrial Europe." *Journal of Development Economics* 127: 339–354.

Jones, Arnold Hugh Martin. 1964. *The Later Roman Empire, 284-602: A Social Economic and Administrative Survey.* Taylor & Francis.

Kessler, David, and Peter Temin. 2008. "Money and Prices in the Early Roman Empire." 137–59. Oxford: Oxford University Press.

Laurence, R. 2002. *The Roads of Roman Italy: Mobility and Cultural Change.* Taylor & Francis.

Leblang, David. 2010. "Familiarity Breeds Investment: Diaspora Networks and International Investment." *American Political Science Review* 104 (3): 584–600.

Litina, Anastasia. 2016. "Natural Land Productivity, Cooperation and Comparative Development." *Journal of Economic Growth* 21 (4): 351–408.

Marichal, Robert. 1988. *Les Graffites De La Graufesenque.* Vol. 47. CNRS.

Masschaele, James. 1993. "Transport Costs in Medieval England." *Economic History Review* 46 (2): 266–279.

McWhirr, Alan. 2002. "Transport by Land and Water." In *The Roman World*, Vol. 2, edited by John Wachter, Chapter 26, 658–670. Routledge.

Mees, Allard W. 2011. *Die Verbreitung von Terra Sigillata aus den Manufakturen von Arezzo, Pisa, Lyon und La Graufesenque: Die Transformation der italischen Sigillata-Herstellung in Gallien.* Verlag des Römisch-Germanischen Zentralmuseums.

Mees, Allard W. 2018. "Was There a Difference between Roman 'civil' and 'military' Samian (terra Sigillata) Market Supply? Finding Answers with Statistical Distribution Analysis Methods." *Internet Archaeology* 50.

Michaels, Guy. 2008. "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System." *Review of Economics and Statistics* 90 (4): 683–701.

Michaels, Guy, and Ferdinand Rauch. 2016. "Resetting the Urban Network: 117–2012." *Economic Journal* 128 (608): 378–412.

Oxé, August, Howard Comfort, and Philip Kenrick. 2000. *Corpus Vasorum Arretinorum: A Catalogue of the Signatures, Shapes and Chronology of Italian Sigillata.* Vol. 2. Bonn: Habelt.

Özak, Ömer. 2018. "Distance to the Pre-industrial Technological Frontier and Economic Development." *Journal of Economic Growth* 23 (2): 175–221.

Pascali, Luigi. 2017. "The Wind of Change: Maritime Technology, Trade, and Economic Development." *American Economic Review* 107 (9): 2821–54.

Paulsson, B. Schulz. 2019. "Radiocarbon Dates and Bayesian Modeling Support Maritime Diffusion Model for Megaliths in Europe." *PNAS* 116 (9): 3460–3465.

Percoco, Marco. 2015. "Highways, Local Economic Structure and Urban Development." *Journal of Economic Geography* 16 (5): 1035–1054.

Persyn, Damiaan, Jorge Díaz-Lanchas, and Javier Barbero. 2020. "Estimating Road Transport Costs Between and Within European Union Regions." *Transport Policy*.

Pétrequin, Pierre, Serge Cassen, Michel Errera, L. Klassen, Alison Sheridan, and Anne Marie Pétrequin. 2012. *Jade: Grandes haches alpines du Néolithique européen, Ve et IVe millénaires av. J.-C.* Centre de Recherche Archéologique de la Vallée de l'Ain.

Polak, Marinus. 1998. "Old Wine in New Bottles. Reflections on the Organization of the Production of Terra Sigillata at La Graufesenque." 115–122. Oxbow Books.

Polak, Marinus. 2000. *South Gaulish Terra Sigillata with Potters' Stamps from Vechten.* Nijmegen: Katholieke Universiteit Nijmegen.

Portes, Richard, and Helene Rey. 2005. "The Determinants of Cross-border Equity Flows." *Journal of International Economics* 65 (2): 269–296.

Rathmann, Michael. 2003. *Untersuchungen zu den Reichsstraßen in den westlichen Provinzen des Imperium Romanum.* Vol. 55. Mainz: von Zabern.

Rauch, James E. 1999. "Networks versus Markets in International Trade." *Journal of International Economics* 48 (1): 7–35.

Rauch, James E., and Vitor Trindade. 2002. "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics* 84 (1): 116–130.

Reddé, Michel. 2018. "The Impact of the German Frontier on the Economic Development of the Countryside of Roman Gaul." *Journal of Roman Archaeology* 31: 131–160.

Redding, Stephen J., and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson and William C. Strange, 1339–1398. Elsevier.

Römisch-Germanisches Zentralmuseum in Mainz. "http://www.rgzm.de/samian/."

Santos Silva, João M.C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics* 88 (4): 641–658.

Santos Silva, João M.C., and Silvana Tenreyro. 2010. "On the Existence of the Maximum Likelihood Estimates in Poisson Regression." *Economics Letters* 107 (2): 310–312.

Schauer, Peter, Andrew Bevan, Stephen Shennan, Kevan Edinborough, Tim Kerig, and Mike Parker Pearson. 2020. "British Neolithic Axehead Distributions and Their Implications." *Journal of Archaeological Method and Theory* 27: 836–859.

Scheidel, Walter. 2014. "The Shape of the Roman World: Modelling Imperial Connectivity." *Journal of Roman Archaeology* 27: 7–32.

Schmidts, Thomas. 2011. *Akteure und Organisation der Handelschifffahrt in den nordwestlichen Provinzen des Römischen Reiches.* Verlag des Römisch-Germanischen Zentralmuseums.

Sidebotham, Steven E. 1986. *Roman Economic Policy in the Erythra Thalassa: 30 B.C.-A.D. 217. Mnemosyne , Vol Suppl. 91*. University Microfilms.

Sotelo, Sebastian. 2019. "Practical Aspects of Implementing the Multinomial PML Estimator." *unpublished manuscript, available for download at http://www-personal.umich.edu/∼ssotelo/.*

Storeygard, Adam. 2016. "Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-saharan Africa." *Review of Economic Studies* 83 (3): 1263–1295.

Tabellini, Guido. 2008. "Institutions and Culture." *Journal of the European Economic Association* 6 (2–3): 255–294.

Talbert, Richard, and Roger Bagnall. 2000. *Barrington Atlas of the Greek and Roman World.* Princeton, N.J.: Princeton University Press.

Temin, Peter. 2012. *The Roman Market Economy.* Princeton University Press.

Wahl, Fabian. 2017. "Does European Development Have Roman Roots? Evidence from the German Limes." *Journal of Economic Growth* 22 (3): 313–349.

Walkowitz, Jürgen. 2003. *Das Megalithsyndrom: europäische Kultplätze der Steinzeit. Beiträge zur Ur- und Frühgeschichte Mitteleuropas*. Beier & Beran.

Whittaker, Charles Richard. 1994. *Frontiers of the Roman Empire: A Social and Economic Study.* Baltimore: Johns Hopkins University Press.

Willis, Steven. 2005. "Samian Pottery, a Resource for the Study of Roman Britain and beyond: The Results of the English Heritage Funded Samian Project. an E-monograph." *Internet Archaeology* 17.

Wilson, Andrew I. 2009. "Large-scale Manufacturing, Standardization, and Trade." In *The Oxford Handbook of Engineering and Technology in the Classical World*, edited by John Peter Oleson, 393–417. Oxford: Oxford University Press.

Wilson, Andrew I., and Alan K. Bowman. 2018. *Trade, Commerce, and the State in the Roman*

World. Oxford University Press.

WISE. 2020. "WISE Water Framework Directive Database."

Wolf, Nikolaus. 2009. "Was Germany Ever United? Evidence from Intra- and International Trade, 1885–1933." *Journal of Economic History* 69 (3): 846–881.

Yeo, Cedric A. 1946. "Land and Sea Transportation in Imperial Italy." *Transactions and Proceedings of the American Philological Association* 77: 221–244.

Yotov, Yoto V., Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. 2016. *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. Geneva, Switzerland, available for download at http://vi.unctad.org/tpa/index.html: United Nations and World Trade Organization.

Appendices

A Main dataset: construction and summary statistics

This appendix describes the construction of the main dataset, including the construction of the Roman transport network, the Roman transport costs, the terra sigillata data, the ownership links data, and all control variables included in the main analysis.

A.1 Constructing the main dataset

The basis of all our datasets are the 903 grid cells within Western Europe that are intersected by the transport network (see Figure A.1). Abstracting from internal trade flows, this means that the total number of potential grid cell pairs is 814,506 (903×903-903).

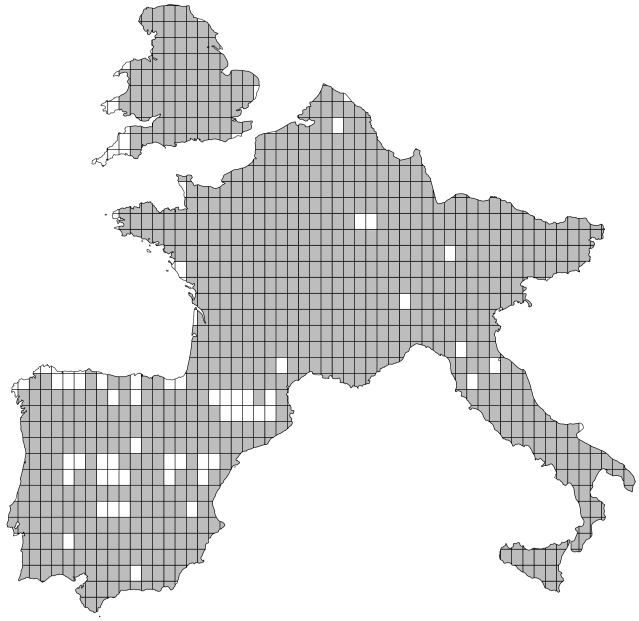


Figure A.1: Extensive Margin of Connectivity to the Roman Transport Network
Figure depicts grid cells intersected by the Roman transport network (shaded grey).

In the context of our analysis it is important to note that whenever a grid cell does not export (or import) at all, the origin (or destination) fixed effect perfectly predicts the outcome. When we observe zero trade flows from a cell to any other cell, the dependent and explanatory variable are collinear. When using the PPML estimator, these perfectly separated observations prevent the estimator from converging. However, the separated observations do not convey relevant information for the estimation process and can thus be safely dropped (Correia, Guimarães and Zylkin, 2020); the Stata command `ppmlhdfe` does this automatically. The implication for our analysis is that depending on the data source, the number of perfectly separated observations, and therefore the size of the estimating dataset, varies. When we focus on the intensive margin of interaction, the size of the dataset depends on the number of grid-cell pairs for which the

dependent variable is greater than zero. Table A.1, lists the different datasets along with the associated number of observations and the estimator for analysis.

Note that, when we use undirected measures of interaction as outcome variables (e.g., time difference in onset of the Plague), we restrict the dataset to unique grid-cell pairs. In this case, the maximum size of the estimating dataset is thus $^{903\times902}/_{2} = 407,253$. Furthermore, we use the OLS estimator instead of the PPML estimator whenever the outcome is a dummy variable (the transformation necessary for estimation with PPML fails due to zeros in these cases) or when values of the outcome can be negative.

A.2 Creating the Roman transport network

For our analysis, we construct a GIS shapefile that represents the Roman transport network. This network spans over the entire Roman empire, as defined at its maximum territorial expansion around 117 CE. The shapefile consists of polylines that trace the routing of Roman roads, rivers that were navigable during Roman times, and coastal routes. The road network is extracted from the digitised version of the Barrington Atlas of the Greek and Roman World (Talbert and Bagnall, 2000). The river network represents river sections that were navigable during Roman times. We compiled this data ourselves (see Table A.2 for details and sources). Transport by sea is possible along the coast. Transshipment between modes is possible at intersections [43]

A.3 Connecting centroids to the network

We connect a grid cell to the Roman transport network by creating an artificial straight-line road segment between its centroid and the closest point on the intersecting network leg(s).[44] On average, the artificial road segments make up 5.7% of the total transport costs. Variation in the length of the added road segment—more precisely the associated costs—represents the fact that average distance to the network varies substantially across grid cells.

Figure A.2 illustrates the connecting procedure for two grid cells—located in central France—with differential access to the network. The lower grid cell is only intersected at the edge. Hence, average distance from points within this cell to the network is large. Conversely, multiple segments of the network cut across the upper grid cell, including one near its centre; average distance to the network is thus much shorter.

It is important to note that our results are not dependent on the choice of how to connect grid cells to the transport network. Estimates are robust to the use of alternative connection procedures. This is illustrated in Tables C.3, where we experiment with per-unit costs of the artificial connections ranging from zero and 200. Similarly, our findings remain unchanged if we incorporate grid cells that are not intersected by the transport network into our analysis (see column (4) in Tables C.1–C.2). In analogy to the procedure outlined above, the non-intersected cells are connected to the network by creating artificial road segments that link the grid cell centroids to the nearest leg of the network .

[43] That is, transshipment is possible where roads intersect with rivers, where roads butt against the coast, and where rivers flow into the sea.

[44] Note that all grid cells in our main analysis are intersected by the network. However, there are substantial differences in the number, location, and type (road, river, sea) of legs that cut across the individual cells. The inclusion of origin and destination fixed effects accounts for such differences.

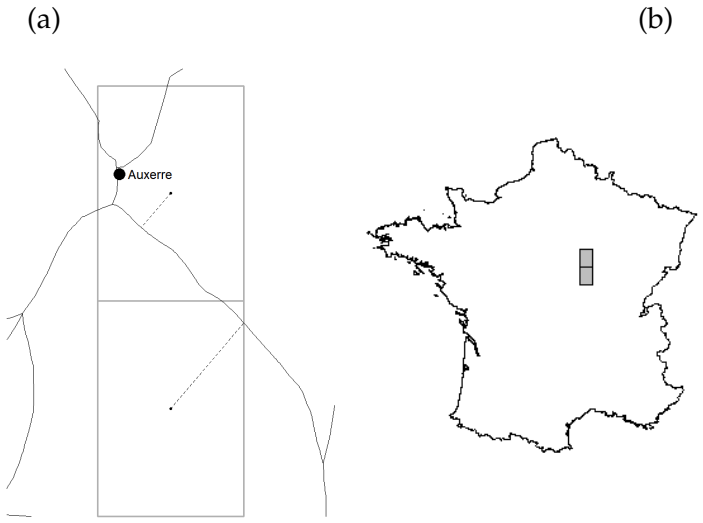


Figure A.2: Connecting Centroids to the Roman Transport Network
Panel (a) depicts two exemplary grid cells in which artificial straight-line road segments (dashed lines) were added to connect centroids to the Roman transport network. Panel (b) shows location of the two grid cells in France.

A.4 Deriving freight rates from Diocletian's Price Edict

Each segment of the transport network has an associated mode of transportation, the cost of which varied substantially across transport modes during Roman times. Our calculations for relative freight rates during the Roman era are based on Diocletian's Price Edict. The Edict was published in 301 CE by the Roman Emperor Diocletian (reign 286–305 CE) and is considered one of the most comprehensive pieces of legislation surviving from antiquity. The edict is divided into one part explaining its intention and a second part listing approximate 1,400 prices for goods and services. These constituted price ceilings that aimed at stabilizing the Roman economy and preventing rent seeking of traders and 'profiteers' (Duncan-Jones, 1982, p. 367). Frequent debasement of the currency by preceding emperors had facilitated inflation and the crisis of the third century during which the empire nearly collapsed. The price ceilings had little impact until the end of Diocletian's reign in 305 CE.

The edict lists the price of land transport by wagon at 20 denarii for transporting 1,200 pounds of wheat per Roman mile, amounting to costs of 0.035 denarii per kg and mile. Downstream river transport is listed at 1 denarius per 20 Roman miles (0.0034 denarii per kg and mile), whereas upstream river transport is listed at 2 denarii (0.0068 denarii per kg and mile). Furthermore, the edict includes freight charges for shipping on 51 sea-routes between specific destinations. Scheidel (2013) lists 48 connections with a clearly identifiable start- and end-point for which he calculates travel distances using the ORBIS: The Stanford Geospatial Network Model of the Roman World (see http://orbis.stanford.edu/?). The resulting distances yield a mean price of 0.00067 denarii per kg of wheat per mile.

Once normalised to the cost of sea travel, average transport costs per kg of wheat per mile are: 1 (sea), 5 (downstream river), 10 (upstream river), 52 (road). By using relative instead of absolute prices, we avoid issues arising from wrongly inferring the price levels as discussed in the literature. Historians generally agree that the relative costs are representative of freight rate differentials during the Roman Empire (Scheidel, 2014; Hopkins, 1983).

A.5 Description of main outcome: terra sigillata

(a)

(b)

(c)

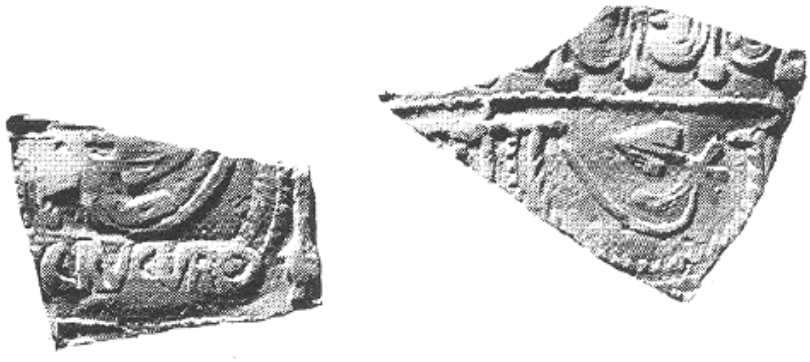


Figure A.3: Production and Design of Roman Terra Sigillata
Panel (a) shows a kiln at La Graufesenque. Panel (b) depicts a variety of terra sigillata products produced at the Rheinzabern site. Panel (c) shows a stamped potsherd that was produced at La Graufesenque and excavated in Kaiseraugst (Switzerland) (source: www.rgzm.de/samian).

To measure bilateral trade volumes during Roman times, we extract information on finds from the terra sigillata database also called Samian navigator (available at www.rgzm.de/samian) which is curated by the Römisch-Germanisches Zentralmuseum in Mainz and is based on the publications of Names on Terra Sigillata (see Hartley et al., 2008) and the Corpus Vasorum Arretinorum (see Oxé, Comfort and Kenrick, 2000). Gallo-Roman terra sigillata is a red-gloss tableware made out of clay which was manufactured at several large production centres in Western Europe. Figure A.3 depicts examples of excavated terra sigillata products and a kiln site. All artefacts included in this database were produced between the beginning of the first century CE and the middle of the third century CE. The database is an ongoing project and is continuously updated. The version of the database used for our analysis was downloaded on November 5, 2020 and includes a total of 242,165 individual artefacts, i.e., potsherds of tableware with known find site and coordinates. The provenance of the artefacts is derived from the potter's stamp which is associated with a production site. Information on the precise location of the production sites can be requested from researchers at the Römisch-Germanisches Zentralmuseum in Mainz.

For the empirical analysis, we exclude (i) artefacts whose find site is located outside of Western Europe, (ii) artefacts that are potential duplicates in the database, (iii) artefacts whose production site is unknown, and (iv) artefacts whose potter stamp cannot be uniquely assigned to a single site.[45] We further exclude (v) artefacts excavated in the same grid cell as the location of their production site. In most of the cases, these finds will represent defect production. Restrictions (i-v) leave us with a total of 107,040 individual artefacts. Table A.3 lists the number of objects included in the analysis by production site.

A.6 Description of main outcome: ownership links

To measure bilateral integration today, we use the number of cross-regional firm ownership links extracted from Bureau van Dijk's Orbis database. We restrict the database to firms with an annual operating revenue of more than 2 million U.S. dollars located within Western Europe. The location of these firms was geocoded manually. Among these firms we extract the subset of firms that are in a cross-regional parent-subsidiary relationship, i.e. firms that either own a stake of at least 25% in another firm located in in a different grid cell or vice versa. The version of the database used for our analysis was downloaded between February–April 2018 and includes 106,996 cross-regional parent-subsidiary links.

A.7 Correlation of distance variables

In the main analysis, we control for a number of least costs paths.

The distance-wise shortest path within the Roman transport network The cost associated with using this path—referred to as network distance—is equal to the length of the path (measured in kilometres). To identify the shortest route, we abstract from any cost differentials across transport modes. That is, we set $\boldsymbol{\alpha} = (1,\ 1,\ 1)$.

The time-wise shortest path within the Roman transport network The cost associated with using this path—referred to as network time—are expressed in hours. To identify the time-

[45] Note that some potters produced at multiple production sites so that their artefacts cannot be assigned to a unique site. However, a number of production sites are also located in close proximity to each other so that they fall into the same grid cell in our dataset. Therefore, in cases when potter stamps are associated with multiple sites but all of them fall into the same grid cell, we include their artefacts in the dataset.

minimising path, we combine the network $\mathbf{N}^{Rome}$ with mode-specific travel speeds $\boldsymbol{\alpha}^{\text{Time}} = (3.7,\ 1.565,\ 2)$. These speeds are measured in km/h and are taken from Carreras and Soto (2013).

The topography-based least-cost path This path is identified on the basis of the Human Mobility Index with Seafaring (HMISea, Özak, 2018). The index is commonly used to proxy for pre-industrial human mobility and takes into account human biological constraints as well as geographical and technological factors. The HMISea least-cost path is not dependent on the transport network $\mathbf{N}^{Rome}$ or $\boldsymbol{\alpha}^{Rome}$. The costs associated with this optimal path is captured by travel time (in hours) and referred to as topography-based distance.

Correlation coefficients for various least-cost measures in the Roman trade sample are displayed in Table 1 in the main text. For the current-day sample they are displayed in Table A.4 below.

A.8 Scatter plots Roman effective distance vs other least-cost measures

Figure A.4 shows the scatterplots between Roman effective distance and various least-cost measures. The fact that the scatter plots in Panels (b)–(c) are bounded from below reflects that the transport-cost minimising path cannot be shorter (quicker) than the distance-minimising (time-minimising) path. The points located on the bounding line represent situations in which the transport cost-minimising path is the same as the distance-minimising (Panel b) or time-minimising (Panel c) path.[46]

A.9 Summary statistics

The main analysis conditions on a number of controls which we define below. Table A.5 reports the summary statistics of all variables in our main datasets.

Joint duration under Roman rule Continuous variable that measures the number of centuries grid cells i and j were jointly part of the Roman territory, as defined in Shepherd (1923).

Absolute distance latitude Continuous variable that measures the absolute difference in decimal degrees between the latitude of the centroids of grid cells i and j.

Same biome Indicator variable that assumes the value one if both grid cells i and j share the same biome, i.e., belong to the same biological community, as defined in Olson et al. (2001).

Access to a waterway Indicator variable that assumes the value one if both grid cells i and j are intersected by a segment of a river or a coastal route, based on our shapefile described above.

Both Mediterranean Sea Indicator variable that assumes the value one if both grid cells i and j touch the Mediterranean Sea.

[46] More precisely, the bounding line represents situations in which the paths are the same and only one mode of transport is used.

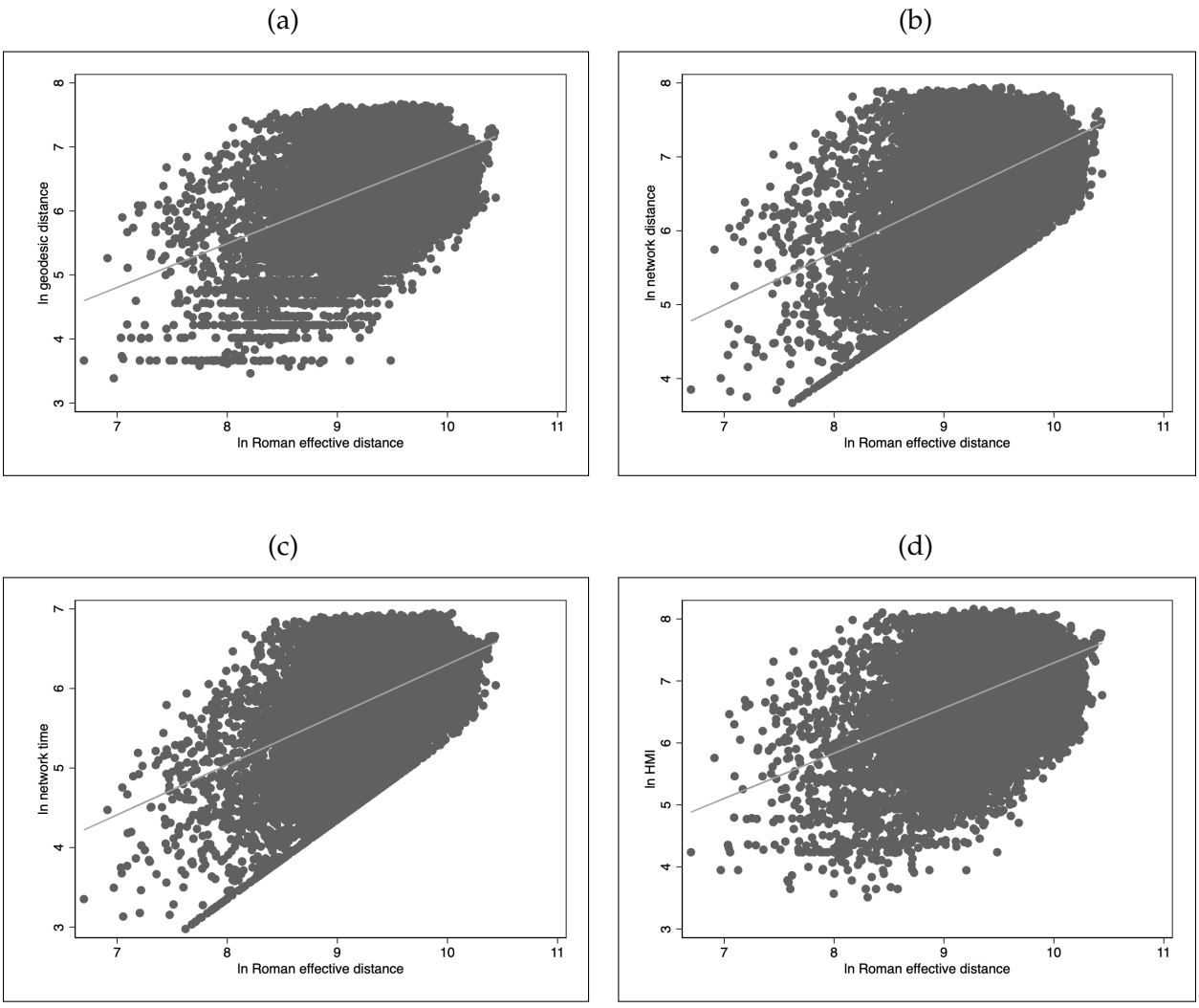


Figure A.4: *Scatterplots Roman effective distance and other least cost measures*
Note. Panel (a) plots Roman effective distance against geodesic distance. Panel (b) plots Roman effective distance against network distance distance. Panel (c) plots Roman effective distance against network time. Panel (a) plots Roman effective distance against the HMISea. All plots also show regression lines that are forced to go through the origin (i.e., the constant is equal to zero).

Table A.1: Estimating datasets and observations

Estimating datasets used in main part and appendices

Estimator	Dataset	Tables used in	# unique origins	# unique destinations	# maximum potential trade pairs	# within trade grids	# actual trade pairs	Comment
PPML	Rome	1, 2, A.5, G.1, C.1	44	520	22,880	41	$44 \times 520 - 41 = 22,839$	
PPML	Orbis	3, A.4, A.5, G.2, C.2	847	865	732,655	832	$847 \times 865 - 832 = 731,823$	
PPML	Orbis (manufacturing)	3 (column (5))	743	812	603,316	719	$743 \times 812 - 719 = 602,597$	
PPML	Orbis (service)	3 (column (6))	662	712	471,344	608	$662 \times 712 - 608 = 470,736$	
PPML	Pre-Roman trade (all goods)	4, E.2	21	378	7,938	15	$21 \times 378 - 15 = 7,923$	
PPML	Pre-Roman trade (Alpine Jade)	4, E.2	7	290	2,300	4	$7 \times 290 - 4 = 2,026$	
PPML	Pre-Roman trade (British Jade)	4, E.2	5	86	430	5	$5 \times 86 - 5 = 425$	
PPML	Pre-Roman trade (Metals)	4, E.2	10	38	380	4	$10 \times 38 - 4 = 376$	
OLS	Pre-Roman cultural diffusion	4, E.3	903	903	815,409	903	$\frac{903 \times 903 - 903}{2} = 407,253$	Undirected measure of interaction.
PPML	Rome IV	5, F.1	38	414	15,732	34	$38 \times 414 - 34 = 15,698$	
PPML	Orbis IV	5, F.1	658	673	442,834	651	$658 \times 673 - 651 = 442,183$	
PPML	Orbis Western Europe Modern Road	6, H.1	844	858	724,152	551	$844 \times 858 - 551 = 723,323$	
PPML	Orbis Non-Roman Empire Modern Roads	6, H.1	574	723	415,002	551	$574 \times 723 - 551 = 414,451$	
PPML & OLS	Orbis Mechanism	7, 8, I.2, I.3	813	831	675,603	798	$813 \times 831 - 798 = 674,805$	

Estimating datasets used in appendices only

Estimator	Dataset	Tables used in	# unique origins	# unique destinations	# maximum potential trade pairs	# within trade grids	# actual trade pairs	Comment
OLS	Price correlation	B.1, B.2					15,159	Undirected measure of interaction. Number of pairs vary across years.
PPML	Plague	B.1, B.2	205	205	42,025	205	$\frac{205 \times 205 - 205}{2} = 20,910$	Undirected measure of interaction.
OLS	Extensive margin pottery	C.1	903	903	815,409	903	$903 \times 903 - 903 = 814,506$	
PPML	Intensive margin pottery	C.1					2,071	Only non-zero flows included.
PPML	Rome (non-connected cells incorporated)	C.1	44	540	23,760	41	$44 \times 540 - 41 = 23,719$	
PPML	Rome (number of pottery finds)	C.5	44	520	22,880	41	$44 \times 520 - 41 = 22,839$	Outcome is count (not share).
OLS	Extensive margin Orbis	C.1	903	903	815,409	903	$903 \times 903 - 903 = 814,506$	
PPML	Intensive margin Orbis	C.2					24,149	Only non-zero flows included.
PPML	Orbis (non-connected cells incorporated)	C.2	918	951	873,018	899	$918 \times 951 - 899 = 872,119$	
PPML	Rome (number of pottery finds)	C.5	44	520	22,880	41	$43 \times 520 - 41 = 22,839$	Outcome is count (not share).
PPML	Orbis (number of ownership links)	C.6	847	865	732,655	832	$847 \times 865 - 832 = 731,823$	Outcome is count (not share).
OLS	Nighttime Light Growth Correlation	D.1, D.2	903	903	815,409	903	$\frac{903 \times 903 - 903}{2} = 407,253$	Undirected measure of interaction.

Notes: Estimating dataset in PPML regressions restricted to observations that are not perfectly separated. In context of PPML, '# unique origins' therefore refers to the number of unique origins that have an outflow greater than zero to at least one destination grid cell. '# unique destinations' refers to the number of unique destination grid cells that have an interaction inflow greater than zero from at least one origin grid cell. Tables are listed in order of first appearance.

Table A.2: Sources for navigable river sections

River	Navigable until:	Source:
Adour	Saint-Sever	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Alb	Ettlingen	Archäologisches Landesmuseum Baden-Württemberg (2006, p. 419)
Allier	Clermont-Ferrand	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Anas	Augusta Emerita	De Soto (2013*a*), Carreras Monfort and De Soto (2009*a*, pp. 303–324), Garcia (1982)
Arno	Arezzo	Campbell (2012, p. 300)
Arroux	near Autun	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Aude (Atax)	Narbonne	Pasquini and Petit (2016, p. 22)
Baetis	Castulo	Munoz (1997, pp. 125–147), De Soto (2013*a*)
Cher	Chateauneuf-sur-Cher	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Colne River	Colchester	Campbell (2012, p. 289)
Deva	fully navigable	De Soto (2013*b*)
Donau	Rißtissen	Archäologisches Landesmuseum Baden-Württemberg (2006, p. 419)
Dordogne	Bergerac	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Doubs	Vesontio	Campbell (2012, p. 69), Moret (2015, Fig 6)
Drava	Klagenfurt	Campbell (2012, p. 292)
Durius	Barca d'Alva (Portugal)	García y Bellido (1944, p. 511), Parodi Álvarez (2012, pp. 137–156)
Ebro	Vareia (Logroño)	Carreras Monfort and De Soto (2009*a*, pp. 303–324), De Soto (2013*b*)
Enz	Pforzheim	Eckoldt (1983, p. 16)
Foss Dyke	Between Trent-River (Torksey) and Witham-River (Lincoln)	Cumberlidge (2009, pp. 120–121), Campbell (2012, p. 289)
Garonne (Garumna)	D'Auterive (just south of Toulouse)	Latour (2006, p. 47)
Inn	Hall	Gattermayr and Steck (2006, p. 7)
Jll	Colmar	Campbell (2012, p. 280), Eckoldt (1986, p. 62)
Limia	fully navigable	García y Bellido (1944, p. 511)
Loir	Chateau-Du-Loir	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Loire (Liger)	Roanne	Williams and Boone (2002, p. 11)
Maenuba	fully navigable	Munoz (1997, pp. 125–147), De Soto (2013*a*)
Main	Mainz	Archäologisches Landesmuseum Baden-Württemberg (2006, p. 388)
Marne	near Saint Dizier	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Mayenne	Mayenne	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Meuse (Mosa)	fully navigable	Wightman (1985, p. 152)
Minius	Lucus Augusti (Lugo)	De Soto (2013*b*)
Mondego River	fully navigable	Parodi Álvarez (2012, pp. 137–156)
Moselle (Mosella)	Epinal	Pasquini and Petit (2016, p. 28)
Nahe	Idar-Oberstein	Dannell and Mees (2015, p. 78)
Neckar	Fischingen	Eckoldt (1983, p. 15)
Nervion	fully navigable	De Soto (2013*b*)
Oise	Tergnier	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Ouse River	York	Campbell (2012, p. 289)
Po	Lago Di Maggiore	Campbell (2012, p. 302)
Rhein	Augusta Raurica (Augst)	Campbell (2012, p. 282)
Rhone	Confluence with Saone at Lugdunum	Campbell (2012, p. 263)
Saar	Saarbrücken	Dannell and Mees (2015, p. 78)
Saone (Arar)	Dijon	Campbell (2012, p. 268)
Sarthe	Le Mans	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Sava	Jesenice	Campbell (2012, p. 292)

continued on next page

River	Navigable until:	Source:
Schelde	Valenciennes	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Segre	from Ebro until Balaguer	Carreras Monfort and De Soto (2008/2009*b*, pp. 313–333)
Seine (Sequana)	near Paris (confluence with Marne)	Campbell (2012, p. 265)
Severn	Gloucester	Campbell (2012, p. 289)
Sil	fully navigable	De Soto (2013*b*)
Tagus	Aranjuey	Carreras Monfort and De Soto (2009*a*, pp. 303–324), García y Bellido (1944, p. 511)
Tarn	near Montauban	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Themse	London	Campbell (2012, p. 289)
Tiber	Città di Castello	Campbell (2012, pp. 309–320)
Trent River	Torksey	Campbell (2012, p. 206 & 289), Cumberlidge (2009, p. 120-121)
Turia	fully navigable	Burriel Alberich, Ribera i Lacomba and Serrano Marco (2004, pp. 129–137)
Vienne	near Limoges	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Villaine	Rennes	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)
Witham River	Lincoln	Cumberlidge (2009, pp. 120–121)
Yonne	Sens	Dannell and Mees (2015, Fig 1), Moret (2015, Fig 6)

Notes: Table displays navigable river sections included in the Roman transport network. Sections are identified as navigable between a river's mouth and a distinct location provided under 'navigable until' if confirmed in the specified source(s).

Table A.3: Production sites, total quantities, and destinations

Production site	# Terra sigillata	# Destinations	Production site	# Terra sigillata	# Destinations
Aachen	2	2	Lezoux	22,385	285
Arezzo	9,764	216	Lot Valley	22	5
Aspiran	5	2	Luxeuil	3	2
Avocourt	150	38	Lyon	952	81
Banassac	1,758	162	Marseille	4	1
Blickweiler	1,166	64	Matres-de-Veyre	5,354	162
Bologna	7	2	Mittelbronn	3	3
Boucheporn	74	18	Montans	1,741	99
Cales	19	10	Nouttre	2	1
Campania	35	7	Pfaffenhofen	34	11
Carrade	5	5	Pisa	3,880	171
Chémery-Faulquemont	609	45	Po Valley	2,953	76
Colchester	15	6	Pont-des-Rames	35	16
Cremona	12	3	Pozzuoli	396	84
Eschweilerhof	13	7	Pulborough	8	3
Espalion	24	13	Rheinzabern	5,220	154
Faenza	3	3	Scoppieto	93	9
Geugnon	53	18	Sinzig	46	14
Haute-Yutz	24	8	Torrita di Siena	26	7
Heiligenberg	655	50	Toulon-sur-Allier	178	28
Ittenwiller	84	18	Trier	1,696	80
Jaulges-Villiers-Vineux	7	5	Valery	153	25
Kräherwald	9	6	Vasanello	95	30
La Graufesenque	44,204	342	Venosa	28	5
La Madeleine	2,279	77	Vichy (Terre-Franche)	25	17
Lavoye	463	60	Vienne	15	10
Le Rozier	141	42	Waiblingen-Beinstein	9	2
Les Allieux	79	26	Westerndorf	25	12

Notes: This table lists terra sigilatta production sites and their excavated output. # *Terra sigillata* represents the total number of terra sigillata finds produced at a given site and excavated within Western Europe, i.e. the geographical scope of our analysis, excluding within production site grid-cell finds. # *Destinations* counts the number of different grid cells in which at least one terra sigillata artefact produced at a given site was found.

Table A.4: Correlation between least cost measures (current-day sample)

	ln Roman effective distance (1)	ln geodesic distance (2)	ln network distance (3)	ln network time (4)
ln geodesic distance	0.293			
ln network distance	0.292	0.979		
ln network time	0.379	0.971	0.979	
ln HMISea	0.332	0.933	0.913	0.909

Notes: This table presents bivariate correlation coefficients between the least-cost measures depicted in Figure 3, based on the current-day sample used in Table 3. 'Roman effective distance' represents the cost associated with shipping goods along the least cost path between grid cells, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'geodesic distance' represents the length in kilometres of the straight-line (as the crow flies) between grid cells. 'network distance' represents the length in kilometres of the distance-wise shortest path between grid cells, given the Roman transport network. 'network time' represents the travel time in hours along the time-wise shortest path between grid cells, given the Roman transport network and Roman-era-specific speed for each mode of transport. 'HMISea' represents the travel time in hours along the time-wise shortest path between grid cells, identified based on the methodology in (Özak, 2018) that incorporates only geographical features and pre-industrial technology.

Table A.5: Descriptive statistics of key variables

Variable	Mean	Std. dev.	Min.	Max.	Obs.
			Roman-era sample		
Share terra sigillata finds	0.023	0.116	0	1	22,839
ln Roman effective distance	9.340	0.442	6.698	10.439	22,839
ln geodesic distance	6.413	0.684	3.386	7.670	22,839
ln network distance	6.668	0.643	3.670	7.941	22,839
ln network time	5.886	0.590	2.977	6.948	22,839
Joint duration under Roman rule	4.397	0.689	3.570	6.720	22,839
Absolute distance latitude	4.244	3.062	0	15.483	22,839
Same biome	0.305	0.461	0	1	22,839
Access to a waterway	0.364	0.481	0	1	22,839
Both Mediterranean Sea	0.011	0.105	0	1	22,839
			Current-day sample		
Share ownership links	0.001	0.015	0	1	731,823
ln Roman effective distance	9.414	0.476	5.953	10.658	731,823
ln geodesic distance	6.640	0.683	3.061	7.802	731,823
ln network distance	6.916	0.648	3.416	8.047	731,823
ln network time	6.093	0.585	2.595	7.116	731,823
Intra-national ownership	0.184	0.388	0	1	731,823
Joint duration under Roman rule	4.410	0.697	3.570	6.720	731,823
Absolute distance latitude	5.101	3.664	0	18.457	731,823
Same biome	0.291	0.454	0	1	731,823
Access to a waterway	0.202	0.402	0	1	731,823
Both Mediterranean Sea	0.016	0.126	0	1	731,823

Notes: This table presents descriptive statistics of variables used for the main analyses in Tables 2 and 3.

B Post-Roman exchange: data and results

This appendix provides empirical support for the qualitative historical evidence that Roman transport network connectivity continued to influenced the intensity of interregional economic exchange after the fall of the Roman Empire. Absent spatially disaggregated and temporally consistent information on trade flows for the post-Roman period, we use price correlation and the time lag in the onset of the Black Death (1346–51) as proxies for market integration. The use of the latter metric is motivated by the fact that the Plague spread along trade routes with merchants being the primary carriers of the disease (see, e.g., Cipolla, 1974; Biraben, 1975; Benedictow, 2006). Differences in the timing of onset can therefore be seen as measure of trade intensity during the Middle Ages (Boerner and Severgnini, 2014).

B.1 Data

Price correlation We compute grid-cell-pair-level price correlations using data compiled in Federico, Schulze and Volckart (forthcoming).[47] Their database reports harmonised information on wheat prices for more than 500 European markets between the 13th and 20th century.[48] For the purposes of our analysis, we use all prices reported for markets located within Western Europe over the time period 1208–1790. This corresponds to the time span between the first year the price data are available and the start of the transportation revolution (i.e., the adoption of the steam engine and the railway). With these restrictions, we are left with 246 markets and 36,477 individual prices. To compute the grid-cell-pair-level price correlation, we first purge the price data of year fixed effects and then calculate the price correlation for each market pair.[49] We then identify into which grid cells the markets fall and compute the average price correlation for each grid cell pair, whereby we weight the individual correlations by the number of observations it is based on. As price correlation is an undirected measure of market integration, we only include unique grid-cell pairs in our analysis. Our final grid-cell-pair-level dataset consists of 15,159 price correlations (see Table A.1 and Table B.1 for key descriptive statistics).

Timing of Black Death onset Differences in the timing of onset of the Black Death are identified based on Christakos et al. (2005, pp. 214–282). This source provides city-level information on the date of the first recorded case of the disease. We manually geocoded the location of all cities for which we can determine the month and year in which the Black Death arrived. For these cities we then identify which grid cell they fall into. Again, we restrict our analysis to Western Europe. This leaves us with 282 cities (located in 205 separate grid cells). The time lag in onset of the epidemic between two grid cells—i and j—is then computed in the following way: First, we determine the time lag in onset of the Black Death (measured in months) for each city pair. In a second step, we compute the average lag for each grid cell pair. Lag in onset of the Plague is, like price correlation, an undirected measure of market integration. We therefore only include unique grid-cell pairs in our analysis. Our final estimating dataset consists of 20,910 observations (see Table A.1 and Table B.1 for key descriptive statistics).

[47] We are grateful to Giovanni Federico for sharing the data.

[48] The data is gathered from great variety of sources, including the well-known Allen-Unger Global Commodity Prices Database.

[49] The price correlation is measured by the Pearson correlation coefficient: $\rho_{m_1,m_2} = \frac{\sum_t (p_{m_1,t} - \bar{p}_{m_1})(p_{m_2,t} - \bar{p}_{m_2})}{(n-1) s_{p_{m_1}} s_{p_{m_2}}}$, where $p_{m_1,t}$ is the wheat price in market m_1 in year t and $p_{m_2,t}$ the price in market m_2 in the same year. $\bar{p}_{m_1}$ and $\bar{p}_{m_2}$ are the average price observed in the respective market. $s_{p_{m_1}}$ and $s_{p_{m_2}}$ is the respective sample standard deviations.

Because the Black Death was introduced to Europe via the port of Messina, differences in connectivity to Messina should predict differences in the timing of the onset of the Plague. The epidemic should (ceteris paribus) reach two regions that are equally well connected to Messina around the same time. On the other hand, timing of onset should differ markedly between areas when connectivity to Messina is very unequal. To test whether the Roman transport network influenced the spatial diffusion of the Black Death, we therefore use a measure that captures differential connectivity to Messina (rather than bilateral connectivity) as explanatory variable. Formally, differential connectivity to Messina between grid cell i and j is defined as:

$$\Delta \text{ effective distance to Messina}_{i,j} = \ln\left(\left|\text{effective distance}_{i,\text{Messina}} - \text{effective distance}_{j,\text{Messina}}\right|\right),$$

where $\text{effective distance}_{i,\text{Messina}}$ is the cost associated with shipping goods from Messina to grid cell i along the least cost path within the Roman transport network. The measure of differential geodesic distance to Messina is defined analogously.[50]

B.2 Results

In Table B.2 we first investigate if price correlation increases with Roman transport network connectivity. Because price correlations range from minus to plus one we use the OLS rather than PPML estimator. Column (1) presents the results. The coefficient of -0.048 implies that price correlation decreases by around 5 percentage points when effective distance increases by one percent.

Column (2) tests if differential connectivity to Messina predicts differences in the timing of the onset of the Plague. This is clearly the case. Greater disparity in connectivity to Messina widens the time lag in the onset of the Black Death between two grid cells. Taken together, the results presented in Table B.2 empirically substantiate qualitative evidence from the historical literature which indicates that the Roman transport network continued to influence the intensity of bilateral trade at least until to the Industrial Revolution.

[50] Formally: Δ geodesic distance to $\text{Messina}_{i,j} = \ln\left(\left|\text{geodesic distance}_{i,\text{Messina}} - \text{geodesic distance}_{j,\text{Messina}}\right|\right)$, where $\text{geodesic distance}_{i,\text{Messina}}$ is geodesic distance between Messina and grid cell i.

Table B.1: Descriptive statistics of price correlation and Plague onset

Variable	Mean	Std. dev.	Min.	Max.	Obs.
	Price Correlation Dataset				
Price correlation	0.190	0.325	-1	1	15,159
ln Roman effective distance	9.283	0.514	6.469	10.496	15,159
ln geodesic distance	6.465	0.685	3.360	7.740	15,159
Joint duration under Roman rule	4.428	0.668	3.570	6.720	15,159
Absolute distance latitude	4.542	3.337	0	16.944	15,159
Same biome	0.306	0.461	0	1	15,159
Access to a waterway	0.322	0.467	0	1	15,159
Both Mediterranean Sea	0.016	0.126	0	1	15,159
	Plague Onset Dataset				
Lag Onset Plague (months)	7.115	5.129	0	26	20,910
ln Δ effective distance to Messina	8.034	1.109	-0.270	9.886	20,910
ln Δ geodesic distance to Messina	12.765	1.164	3.245	14.621	20,910
Joint duration under Roman rule	4.235	0.686	3.570	6.720	20,910
Absolute distance latitude	4.781	3.490	0	16.893	20,910
Same biome	0.294	0.455	0	1	20,910
Access to a waterway	0.205	0.403	0	1	20,910
Both Mediterranean Sea	0.016	0.124	0	1	20,910

Notes: This table presents descriptive statistics of the main dependent and explanatory variables used in Table B.2

Table B.2: Roman transport network connectivity and economic integration 1208–1790

Dependent Variable:	Price Correlation	Lag Onset Plague
	(1)	(2)
ln Roman effective distance	-0.046**	
	(0.022)	
ln geodesic distance	-0.209***	
	(0.012)	
ln Δ Roman effective distance to Messina		0.033***
		(0.011)
ln Δ geodesic distance to Messina		0.020
		(0.018)
Baseline controls	Yes	Yes
Destination FEs	Yes	Yes
Origin FEs	Yes	Yes
Observations	15,159	20,910
Estimator	OLS	PPML

Notes: This table reports estimates of Equation (1) using the OLS estimator (column 1) and PPML estimator (column 2). Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. The dependent variable in column (1) is the grid-cell-pair-level correlation of wheat prices between grid cells i and j compiled in Federico, Schulze and Volckart (forthcoming). The dependent variable in column (2) is the difference in the timing of onset of the Black Death between grid cells i and j in months. The dependent variables are described in detail in Appendix B. Baseline controls correspond to column 2 in Table 2 and are described in the respective table notes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Robustness tests

This appendix presents results of the robustness checks discussed in Section 5.3, in which we vary definitions and assumptions made when constructing and analysing the main datasets.

Varying definitions and assumptions Table C.1 presents data on trade during the Roman era, whereas Table C.2 presents data on interregional firm ownership today. In both tables, column (1) estimates the extensive margin of interaction using Equation (1) in a linear probability model. That is, the dependent variable is an indicator that assumes the value one if grid cells i and j have any form of interaction in the respective period. It therefore uses the full sample of grid-cell pairs. In both tables, column (2) estimates the intensive margin of interactions using Equation (1) in a PPML model. That is, the dependent variable is the share of interactions in grid cell i originating from grid cell j in the respective period using only positive interaction flows. Results confirm that Roman effective distance is related to the intensive as well as the extensive margin of integration.

In both tables, column (3) presents results from estimating the baseline model including grid cells that are not intersected by the Roman transport network. In analogy to our standard procedure, these cells are connected to the network by creating artificial road segments that link the grid cell centroids to the nearest leg of the network.

In both tables, column (4) presents results from estimating the baseline model assuming differential costs depending on the direction of trade. In the main analysis, we abstract from the fact that downstream-river transport is less costly than upstream-river transport. For this robustness test, we assume mode-dependent relative costs of maritime transport (1), downstream-river transport (5), upstream-river transport (10), and road transport (52), and estimate the model using the resulting measure of Roman effective distance. In both tables, column (5) presents results from estimating the baseline mode using an alternative way to cluster standard errors (clustering at 1×1 degree grid cells).

Table C.2, column (6) presents results from estimating the baseline model including a full set of country-pair fixed effects in addition to the origin and destination fixed effects. Finally, column (7) presents results from estimating the baseline model when changing the cut-off in the ownership definition. Instead of defining ownership as having a minimum stake of 25% in another firm, it defines ownership as having a minimum stake of 50%.

Table C.3, documents that out results do not depend on the exact choice of per-unit transport costs of the artificial roads (see Appendix A.3 for details).. Varying the (relative) cost of the artificial roads between zero and 200 has little impact on our estimates.

Production-site-level analysis For the main analysis, we aggregate Roman trade across production sites within grid cells. In Table C.4, we estimate the standard model at the production-site-destination-grid-cell level. In these regressions, model (1) to includes production-site and destination grid cell fixed effects.[51] The results are almost identical.

Counts In Tables C.5–C.6 we use counts rather than shares as outcome variables. That is, in C.5 the dependent variable is the number of potter finds excavated in grid j that originates from grid i. In Table C.6 the outcome is the number of firms located in grid j that are part owned by firms located in grid i.

[51] Our main model controls for origin grid cell and destination grid cell fixed effects.

Table C.1: Robustness tests–Roman era

Dependent Variable:	Share of Terra Sigillata Finds				
	(1)	(2)	(3)	(4)	(5)
ln Roman effective distance	-0.010***	-0.489**	-1.414**	-1.341***	-1.493***
	(0.003)	(0.218)	(0.550)	(0.468)	(0.549)
ln geodesic distance	-0.001	-0.727***	-0.721***	-0.694***	-0.655***
	(0.001)	(0.187)	(0.230)	(0.224)	(0.232)
Robustness	Extensive margin	Intensive margin	Include non-connected grids	Directed freight rates	Clustering
Baseline controls	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes
Observations	814,506	2,071	23,719	22,839	22,839
Estimator	OLS	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1). Column (1) estimated by OLS; columns (2)–(5) by PPML. Dependent variable is the share of terra sigillata finds in a cell j that originates from cell i. Column (1) estimates a linear probability model, where the dependent variable is an indicator that is one if there is any exchange of terra sigilatta between grid cells i and j. Column (2) estimates the model in a sample of grid-cell pairs for which we observe at least one terra sigilatta exchange between grid cell i and j. Column (3) estimates the model in a sample of all grid cells, including cells that are not crossed by the transport network. Column (4) uses a measure of Roman effective distance that imposes differential cost for up- and downstream river transport. In columns (1–4), standard errors, reported in parentheses, are two-way clustered at the origin and destination grid-cell level. In column (5), standard errors, reported in parentheses, are two-way clustered at a level that combines grid cells into "super grid cells" of 1×1 degree size. Baseline controls correspond to column 2 in Table 2 and are described in the respective table notes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.2: Robustness tests–Today

Dependent Variable:	Share of Ownership Links (>25% Ownership)						Share of Ownership Links (>50% Ownership)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln Roman effective distance	-0.081***	-0.251***	-0.370***	-0.394***	-0.404***	-0.382***	-0.354***
	(0.006)	(0.055)	(0.088)	(0.077)	(0.091)	(0.073)	(0.085)
ln geodesic distance	-0.044***	-0.657***	-1.490***	-1.519***	-1.521***	-1.572***	-1.397***
	(0.003)	(0.039)	(0.057)	(0.059)	(0.077)	(0.059)	(0.063)
Robustness	Extensive margin	Intensive margin	Include non-connected grids	Directed freight rates	Clustering	Country-Pair FE	Cut-off 50% ownership
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	814,506	24,149	872,111	731,823	731,823	731,823	697,731
Estimator	OLS	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1). Column (1) estimated by OLS; columns (2)–(7) by PPML. Dependent variable is the share of firms in cell j that are (partly) owned by firms located in cell i. Column (1) estimates a linear probability model, where the dependent variable is an indicator that is one if there is any exchange of terra sigilatta between grid cells i and j. Column (2) estimates the model in a sample of grid-cell pairs for which we observe at least one terra sigilatta exchange between grid cell i and j. Column (3) estimates the model in a sample of all grid cells, including cells that are not crossed by the transport network. Column (4) uses a measure of Roman effective distance that imposes differential cost for up- and downstream river transport. In columns (1–4) and (6–7), standard errors, reported in parentheses, are two-way clustered at the origin and destination grid-cell level. In column (5), standard errors, reported in parentheses, are two-way clustered at a level that combines grid cells into "super grid cells" of 1×1 degree size. Column (6) adds a full set of country-pair fixed effects. Column (7) changes the threshold for the definition ownership in the dependent variable to a minimum of 50%. Baseline controls correspond to column 2 in Table 3 and are described in the notes of Tables 2 and 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Varying cost of artificial roads segments

Dependent Variable:	Share Pottery Finds					Share Ownership Links (>25% Ownership)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln Roman effective distance	-1.280***	-1.493***	-1.623***	-1.720***	-1.790***	-0.319***	-0.404***	-0.431***	-0.446***	-0.456***
	(0.460)	(0.542)	(0.587)	(0.621)	(0.645)	(0.060)	(0.077)	(0.087)	(0.095)	(0.101)
ln geodesic distance	-0.675***	-0.655***	-0.657***	-0.667***	-0.682***	-1.534***	-1.521***	-1.528***	-1.538***	-1.547***
	(0.235)	(0.230)	(0.226)	(0.223)	(0.220)	(0.060)	(0.060)	(0.061)	(0.061)	(0.061)
Same country	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Joint duration under Roman rule	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,839	22,839	22,839	22,839	22,839	731,823	731,823	731,823	731,823	731,823
Cost artificial roads	0	52	100	150	200	0	52	100	150	200
Estimator	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Dependent variables are the share of terra sigillata finds in cell i that originates from cell j (columns 1–3) or the share of firms in cell i that are (partly) owned by firms located in grid j (columns 4–6). 'Roman effective distance' represents the cost associated with shipping goods along the least cost path, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. 'geodesic distance' represents the length in kilometres of the straight-line (as the crow flies) connection between grid cell centroids. Further controls are described in notes of Table 2 and and 3 in the main text. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Further robustness tests–Production-site-level analysis

Dependent Variable:	Share of Terra Sigillata Finds					
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	-2.895***	-2.100***	-2.053***	-1.493***	-1.498***	-1.254***
	(0.564)	(0.541)	(0.501)	(0.541)	(0.521)	(0.478)
Joint duration under		2.943***	2.638***	2.277***	2.277***	2.287***
Roman rule (centuries)		(0.287)	(0.367)	(0.277)	(0.277)	(0.263)
ln geodesic distance				-0.655***	-0.679	-0.038
				(0.232)	(0.639)	(0.524)
ln network distance					0.029	
					(0.647)	
ln network time						-0.886
						(0.568)
Geography controls	No	No	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Production Site FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	29,067	29,067	29,067	29,067	29,067	29,067
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates from an adjusted version of Equation (1) using the PPML estimator. Instead of aggregating production sites to the grid-cell level, in this table, origins are defined as production sites. Equation (1) is therefore adjusted to include production-site fixed effects instead of grid-cell fixed effects. Standard errors two-way clustered at the production site and destination grid cell level are reported in parentheses. Dependent variable is the share of terra sigillata finds in cell i that originates from production site j. For details on the explanatory variables, see notes of Tables 2 in the main text. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: Further robustness tests–Using counts instead of shares–Roman era

Dependent Variable:	Number of Terra Sigillata Finds					
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	-3.011***	-2.158***	-2.545***	-1.388***	-1.545***	-1.528***
	(0.550)	(0.472)	(0.455)	(0.413)	(0.446)	(0.441)
Joint duration under		3.705***	3.376***	2.995***	2.976***	3.024***
Roman rule (centuries)		(0.494)	(0.496)	(0.533)	(0.493)	(0.538)
ln geodesic distance				-1.277***	-2.646***	-1.816**
				(0.341)	(0.899)	(0.867)
ln network distance					1.562	
					(0.956)	
ln network time						0.679
						(0.960)
Geography controls	No	No	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,839	22,839	22,839	22,839	22,839	22,839
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid-cell level are reported in parentheses. Dependent variable is the count (instead of the share) of terra sigillata finds in cell j that originates from cell i. For details on the explanatory variables, see notes of Tables 2 in the main text. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.6: Further robustness tests–Using counts instead of shares–Today

Dependent Variable:	Number of Ownership Links (>25% Ownership)					
	Full Sample				Manufacturing	Service
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	-0.533***	-0.405***	-0.430***	-0.519***	-0.421***	-0.600***
	(0.103)	(0.110)	(0.117)	(0.122)	(0.106)	(0.183)
ln geodesic distance	-1.214***	-1.407***	-1.480***	-1.621***	-1.456***	-0.926***
	(0.073)	(0.088)	(0.145)	(0.139)	(0.083)	(0.145)
Intra-national ownership	1.224***	1.123***	1.129***	1.136***	1.019***	1.565***
	(0.113)	(0.129)	(0.129)	(0.129)	(0.118)	(0.149)
Joint duration under Roman rule (centuries)		0.576***	0.576***	0.581***	0.624***	0.350
		(0.162)	(0.162)	(0.160)	(0.154)	(0.230)
ln network distance			0.100			
			(0.144)			
ln network time				0.329**		
				(0.164)		
Geography controls	No	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	731,823	731,823	731,823	731,823	602,597	470,736
Estimator	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid-cell level are reported in parentheses. Dependent variable is the count (instead of the share) of firms in cell j that are (partly) owned by firms located in cell i. For details on the explanatory variables, see notes of Tables 2 and 3 in the main text. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Business cycle co-movement: data and results

To investigate if the degree of business-cycle integration is influenced by connectivity within the Roman transport network, we draw on night-time luminosity data from the Defense Meteorological Satellite Program-Optical Line Scanner (DMSP-OLS) sensor. This data is available for the years 1992–2013 at a spatial resolution of 1×1 kilometres. Based on this information, we first determine overall night-time light intensity for each grid cell and year by summing up the light intensity indices of the individual 1×1 km pixels that fall into a given 0.5×0.5 degree grid cell. We then compute the annual growth rates between 1992 and 2013 for each grid cell. In a final step, we create two proxies for business cycle co-movement: The first is defined as the simple correlation coefficient in night-time light growth; the second as the correlation coefficient after the cyclical component has been removed using the Baxter-King filter. Table D.1 reports summary statistics of the dataset.

Table D.2 presents the estimates obtained from regressing proxies for business cycle integration on Roman effective distance. For both measures we find that greater connectivity within the Roman transport network increases business cycle synchronisation. The coefficients imply that correlation in night-time light growth decreases by 1.3–3.1 percentage points when Roman effective distance increases by 1%. The fact that Roman effective distance affects both the intensity of interregional firm links and business cycle transmission accords with recent cross-country evidence on the close interrelationship between these two aspects of economic integration (Cravino and Levchenko, 2017).

Table D.1: Descriptive statistics business cycle

Variable	Mean	Std. dev.	Min.	Max.	Obs.
Nighttime Light Growth Correlation (no filter)	0.651	0.165	-0.563	0.991	407,253
Nighttime Light Growth Correlation (filter)	0.757	0.137	-0.347	0.996	407,253
ln Roman effective distance	9.414	0.481	5.628	10.658	407,253
ln geodesic distance	6.641 0	0.685	2.518	7.802	407,253

Notes: This table presents descriptive statistics of variables used for the extension to business cycle co-movement analysis in Table D.2.

Table D.2: Roman transport network connectivity and business cycle integration

Dependent Variable:	Nighttime Light Growth Correlation	
	(1)	(2)
ln Roman effective distance	-0.013***	-0.031***
	(0.004)	(0.004)
ln geodesic distance	-0.051***	-0.019***
	(0.003)	(0.002)
Intra-national ownership	-0.004*	-0.003
	(0.002)	(0.002)
Joint duration under Roman rule (centuries)	0.007**	0.018***
	(0.003)	(0.003)
Mean of dep. var.	0.651	0.757
SD of dep. var.	0.165	0.137
Geography controls	Yes	Yes
Baxter-King filter	no	yes
Destination FEs	Yes	Yes
Origin FEs	Yes	Yes
Observations	407,253	407,253
Estimator	OLS	OLS

Notes: This table reports estimates of Equation (1) using the OLS estimator. Standard errors two-way clustered at the origin and destination grid-cell level are reported in parentheses. Dependent variable is the correlation coefficient of nighttime-light growth rates during 1992–2013 between grid cells i and j. This is an undirected measure, therefore only unique grid cell pairings are included in the regression. Column (2) adds the Baxter-King filter to remove the cyclical component in the dependent variable. For details on the explanatory variables, see notes of Tables 2 and 3 in the main text. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E Pre-Roman exchange: data, sources, and summary statistics

This appendix describes the data collection and structure of datasets used to test the relationship of Roman connectivity and interaction prior to the Roman era in Section 6.

E.1 Bilateral Trade in Goods

To test whether Roman transport network connectivity influences the movement of goods prior to the existence of the network, we systematically scoured the archaeological literature for data on the intensity of pre-Roman exchange of goods. A prerequisite for the data to be included in our analysis is that the precise origin (provenance) and destination (find site) of the goods could be inferred. Our extensive search yielded the following results: we identified two existing large-scale databases on trade during Neolithic time that report precise origin and destination of artefacts. Additionally, we compiled our own dataset on Bronze age trade by collating information from a large variety of separate publications that fulfil the inclusion criterion. The characteristics of the three databases—the two existing ones and our own collection—are described below.

Alpine jade–Neolithic period The Alpine jade database was compiled by more than 50 researchers from several European countries as part of the project "JADE: Social inequalities in Neolithic Europe: the circulation of long axeheads of Alpine jades" (Pétrequin et al., 2012). The data was compiled between 2008 and 2018 and focuses on long axeheads made from jade (mainly extracted in the Alps). In total, the database contains precise information on the find site of 2,173 jade axeheads which can be downloaded at http://jade.univ-fcomte.fr. These axes mainly circulated within Europe during the 5th and 4th millennia BCE, but some of the objects moved over long distances from the Alpes to the Atlantic coast and the Black Sea via extensive exchange networks. To identify the provenance of the axeheads, the researchers used visual analysis and spectroradiometry. For 1,355 axeheads, the database includes information on the origin of the artefacts with sufficient precision for the purpose of our analysis. The jade included in our analysis primarily originates from high-altitude quarries at Monte Viso near Turin (973 pieces) and Monte Beigua near Genoa (151 pieces) in the Italian Alps. Some of the finds can also be reliably traced back to the Vosges in France or the Pennine Alps in Switzerland.

British axeheads–Early Neolithic period Schauer et al. (2020) provide an extensive database on British axeheads that combines information collected in projects overseen by the Implement Petrology Group (IPG), the Neolithic Axehead Archive, and the Irish Stone Axe Project. The database contains precise information on the find site of 5,809 axeheads from the Early Neolithic (4100–3400 BCE) discovered across England, Wales and southern Scotland. Provenance of the axeheads is identified via petrological analysis. For our analysis, we exclude axeheads made from flint (1,512 pieces) due to fact that origins could not be unambiguously determined. We further exclude axeheads for which the source was specified as 'other' (1,766 pieces) which includes jade pieces originating from continental Europe. These restrictions leave us 2,345 artefacts for which the provenance is pinpointed within a radius of less than 50 km in Schauer et al. (2020).

Metals–Bronze age This database is the result of our own extensive literature search. It combines information reported in 36 separate publications (listed in Table E.1), which predominantly describe objects traded during the Bronze age. A large share of the objects are weapons, tools,

and jewellery produced between 1500 BCE and 1100 BCE.[52] In total, our database includes 3,744 artefacts for which find site and provenance could be determined with sufficient precision.[53] Based on information reported in the original publications, we georeferenced the find site and provenance of each artefact. The exact site or the municipality of excavation was specified in 90 percent of cases; for the remaining artefacts, the destination was identified based on the region or province of the find. In these cases we define the centroid of the corresponding area as the destination.

In most cases, the publications identify an artefact's provenance based on its metal parts, such as copper, silver, tin, or lead, using lead isotope analysis and trace element pattern analysis.[54] In few cases it is determined by typology of similar instruments. The combination of these methods allows researchers to reliably relate metal artefacts to specific ore deposits from which they originate.[55] For 385 artefacts in our database, the origin of their amber parts was identified. The precise site of provenance is reported for around half of the artefacts. In the remaining cases, the provenance is identifiable based on geographical features, often a mountain or a specific part of a mountain range. In these cases, we geocoded the provenance using the centroid of the mountain range.

Table E.1 lists the original sources along with the number of artefacts incorporated in our analysis. Figure E.1 depicts the origins (Panel a) and destinations (Panel b) of the artefacts. Table E.2 presents the summary statistics of the datasets.

Aggregation In keeping with the main analysis, we aggregate the information of the individual artefacts to the grid-cell-pair level. To this end, we first identify in which grid cell its origin and destination lies. We then aggregate this information to the grid-cell-pair level giving us the number of finds within cell j that originate from cell i. Again, we follow Eaton, Kortum and Sotelo (2013) and define the share of j's total imports that originate from i as our measure of interregional trade flows. Table E.2 reports descriptive statistics separately for the three datasets (Panels b–d). Panel (a) shows the characteristics of the combined dataset (created by treating any artefact—irrespective of its source—as one trade flow).

In Table 4 columns 1–5 we focus on artefacts for which both, origin and destination are located within Western Europe. To increase the statistical power of our analysis, we extend our dataset to include finds for which only the origin must fall into Western Europe, whereas the destination may be located outside of our regions of analysis. To integrate these observations into the Roman transport network, we have to make a somewhat arbitrary decision about how to connect the destinations that lie outside Western Europe to the Roman transport network. For simplicity and transparency, we decided to assign each destination to the nearest grid cell (measured by the geodesic distance) within Western Europe. To account for the fact that the distance

[52] The predominant view in the literature is that during the Bronze age, pottery and earthenware goods were traded only very locally. However, according to Sabatini and Lo Schiavo (2020), pottery was sometimes transported alongside more valuable goods. Their database includes 14 ingots and artefacts from copper but also 37 pieces of pottery made from clay that were traded in the same context.

[53] Among the 3,744 artefacts, only 786 item have both find site *and* provenance in Western Europe. Thus, in an extension of our baseline analysis described below, we develop a methodology to include those artefacts whose provenance is located in Western Europe but whose find site is located outside of Western Europe into our analysis.

[54] Note that an unknown number of artefacts in our database consists of multiple metal parts that may originate from different deposits. Thus, the same artefact may be subject to 'provenancing' of its copper parts in one study and 'provenancing' of its silver parts in another.

[55] Lead isotope analysis relies on the fact that lead isotopes continuously change until the ore deposit is formed. Deposits form at specific geological times so that isotope ratios from metal samples and deposits can be compared and the most likely provenance can be determined. For more information, see Radivojević et al. (2019).

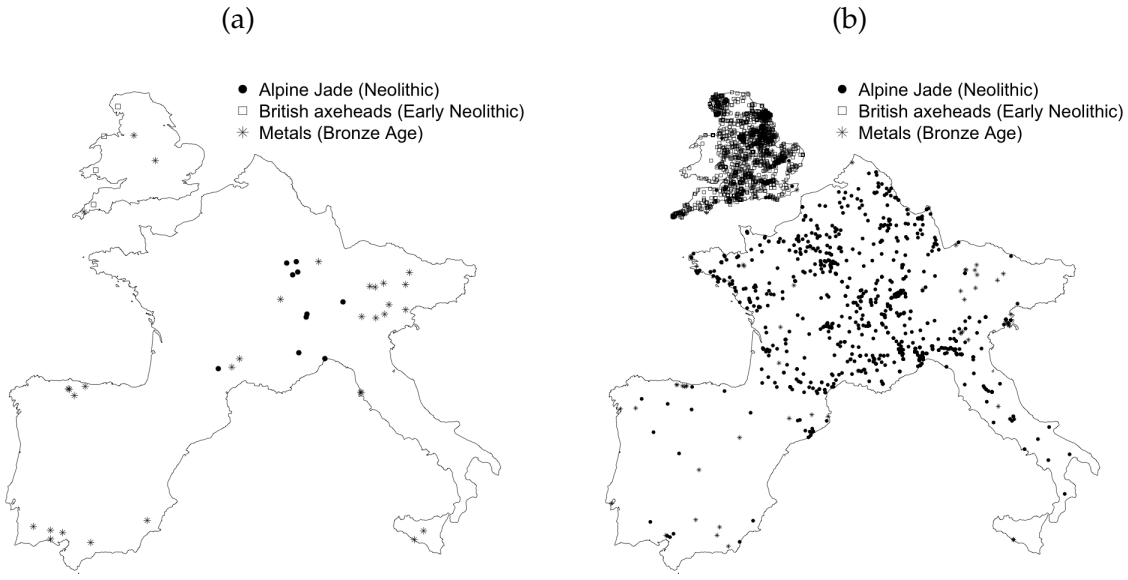


Figure E.1: Origins and Destinations of Pre-Roman Artefacts
Panel (a) depicts origins; panel (b) shows the spatial distribution of excavation sites. Figure is restricted to geographical scope of our analysis.

to the Western European border varies across excavation sites (i.e., destinations), we weight the finds using the inverse distance.[56] Not employing weights produces very similar results. The incorporation of the additional destinations almost doubles our sample size. Descriptive statistics of this extended dataset are described in Panel (e) of Table E.2.

E.2 Diffusion of culture

As outlined in Section 6, a generally accepted view among archaeologists is that the spatiotemporal diffusion of burial traditions took place by way of cultural exchange, including migration (Cummings, Midgley and Scarre, 2015, p. 825 ff., Paulsson, 2019, Holst, 2013, p. 117, Childe, 1958, p. 123 ff., Childe, 1930, p.173 ff.).[57] The presence of the same type of burial site in two regions therefore implies economic and social interaction. We collected information on the location of remains for such burial traditions for periods of human history prior to the advent of Rome. Table E.3 presents the summary statistics of the dataset.

Burial traditions–Neolithic period As proxy for interaction during the Neolithic, we use the common presence of megalithic structures (megalithic graves such as dolmen and passage graves, called chamber cairns), the dominant standardised burial tradition of this time (Cummings, Midgley and Scarre, 2015). Information on the location of dolmen and chambered cairns is derived from the Megalithic Portal.[58] Furthermore, we collected information on the location

[56] Formally, the weight w_i is given by: $w_i = 1 - d_i / d^{\max}$, where i is a given find and d_i its distance to Western Europe. For finds excavated within Western Europe, this distance is zero. These observations thus receive a weight of one. $d^{\max}$ is the maximum distance between an excavation site that lies outside of Western Europe and the Western European border.

[57] See Holst (2013) for a discussion.

[58] For details see http://bit.do/burialchambers and http://bit.do/chamberedcairns.

of menhirs, also from the Megalithic Portal.[59] Menhirs are standing stones which are associated with cultural and religious rituals in the late Neolithic and early Bronze age in Europe (Walkowitz, 2003).

Burial traditions–Bronze age Interregional interaction during the Bronze Age is measured by the common occurrence of the most widespread burial ritual, the Tumulus tradition (e.g., round barrows) which emerged rather suddenly in large parts across Europe and is associated with elite formation (Holst, 2013, p. 103, Darvill, 2013, p. 144). Information on the location of round barrows is derived from the Megalithic Portal.[60]

Settlements–Iron age In addition to these indicators, we employ the (concurrent) presence of Celtic settlements (Oppida) during the La Tène culture in Gaul in grid-cell pairs as a proxy for (socio-)economic interaction. This is motivated by the fact that the Celtic culture spread across Europe via migration in the Iron Age. Information on the location of Oppidum settlements is derived from the Oppida portal, an initiative of the Marc Bloch University.[61]

We define the grid-cell-pair level measure of interaction as an indicator variable that takes the value of one if the same type of burial site (or oppidum) is found in cell i and cell j.

[59] For details see http://bit.do/menhirs.
[60] For details see http://bit.do/roundbarrows.
[61] For details see http://bit.do/oppida.

Table E.1: References for database of Bronze Age metals exchange

Commodity	# Items	# Destinations	# Origins	Source
Bronze axes	2	2	1	Artioli et al. (2020)
Amber jewelry and ornaments	46	1	1	Beck, Fellows and Adams (1970)
Amber jewelry	70	1	1	Beck, Southard and Adams (1972)
Copper ingots	47	6	1	Begemann et al. (2001)
Bronze weapons	9	8	5	Burgess and O'Connor (2008)
Bronze axes and copper ingots	17	2	3	Canovaro et al. (2019)
Lead utensils	71	12	2	Gale and Stos-Gale (1981)
Lead utensils	40	1	2	Gale, Stos-Gale and Davis (1984)
Copper ingots and lead artefacts	614	7	3	Gale and Stos-Gale (2005)
Copper ingots	79	4	1	Gale (2006)
Copper artefacts	120	1	1	Giumlia-Mair (2005)
Copper artefacts	24	2	1	Hauptmann, Begemann and Schmitt-Strecker (1999)
Copper ingots	8	4	1	Kaiser (2013)
Bronze axes	3	3	2	Klassen, Cassen and Pétrequin (2012)
Copper ingots	4	1	1	Kristiansen and Suchowska-Ducke (2015)
Bronze weapons	53	32	12	Ling et al. (2014)
Bronze swords	108	78	8	Ling et al. (2019)
Copper ingots and ceramic goods	3	3	2	Lo Schiavo (2005)
Tin artefacts	142	63	2	Mason (2020)
Bronze utensils	1227	21	2	Matthäus (2005)
Bronze weapons and jewelry	69	38	10	Melheim et al. (2018)
Bronze weapons	1	1	1	Montero Ruiz, Martínez Navarrete and Galán (2016)
Bronze weapons and copper utensils	13	1	3	Montero Ruiz et al. (2018)
Amber jewelry	1	1	1	Murillo-Barroso and Martinón-Torres (2012)
Amber jewelry	268	5	1	Murillo-Barroso et al. (2018)
Bronze weapons and utensils	19	19	1	Needham and Giardino (2008)
Bronze weapons	383	7	2	Pernicka, Lutz and Stöllner (2016)
Bronze weapons and jewelry	12	1	1	Pernicka et al. (2016)
Copper ingots	8	1	2	Pinarelli (2004)
Bronze axes and copper ingots	27	4	3	Reguera-Galan et al. (2019)
Bronze weapons and jewelry	3	1	3	Reiter et al. (2019)
Copper ingots and ceramic goods	51	8	9	Sabatini and Lo Schiavo (2020)
Silver coins	12	1	1	Stos-Gale and Gale (2009)
Bronze, tin and silver artefacts	99	21	7	Stos-Gale and Gale (2010)
Copper ingots	17	3	1	Stos-Gale (2011)
Bronze weapons	74	3	1	Williams and Le Carlier de Veslud (2019)

Notes: This table lists the number, destination, and origins of artefacts by reference. These data were collected to compile the database of metal artefacts from the Bronze age used for estimation in Table 4.

Table E.2: Descriptive statistics of pre-Roman trade flows

Variable	Mean	Std. dev.	Min.	Max.	Obs.
			Panel (a): All databases		
Share trade	0.048	0.197	0	1	7,923
ln Roman effective distance	9.416	0.514	6.639	10.311	7,923
ln geodesic distance	6.482	0.679	3.371	7.753	7,923
			Panel (b): Alpine jade–Neolithic		
Share trade	0.143	0.338	0	1	2,026
ln Roman effective distance	9.569	0.422	7.289	10.311	2,026
ln geodesic distance	6.101	0.627	3.536	7.396	2,026
			Panel (c): British axeheads–Early Neolithic		
Share trade	0.203	0.288	0	1	425
ln Roman effective distance	8.478	0.491	6.639	9.425	425
ln geodesic distance	5.479	0.558	3.371	6.272	425
			Panel (d): Metals–Bronze age		
Share trade	0.101	0.297	0	1	376
ln Roman effective distance	9.477	0.466	7.818	10.243	376
ln geodesic distance	6.601	0.805	3.789	7.608	376
			Panel (e): All databases, origins outside Western Europe included		
Share trade	0.026	0.147	0	1	15,143
ln Roman effective distance	9.449	0.491	6.639	10.369	15,143
ln geodesic distance	6.599	0.688	3.371	7.776	15,143

Notes: This table presents descriptive statistics of variables used for the analyses of pre-Roman trade in goods in Tables 4 Panel A.

Table E.3: Descriptive statistics of pre-Roman cultural diffusion

Variable	Mean	Std. dev.	Min.	Max.	Obs.
Both Cells Dolmen	0.135	0.342	0	1	407,523
Both Cells Chambered Cairns	0.002	0.047	0	1	407,523
Both Cells Menhir	0.159	0.365	0	1	407,523
Both Cells Round Barrow	0.30	0.170	0	1	407,523
Both Cells Oppida	0.014	0.116	0	1	407,523
ln Roman effective distance	9.414	0.481	5.628	10.658	407,523
ln geodesic distance	6.661	0.685	2.518	7.803	407,523

Notes: This table presents descriptive statistics of variables used for the analyses of pre-Roman diffusion of culture in Tables 4 Panel B.

F IV approach: construction and robustness

This appendix provides a detailed description of how we construct the instrumental variable to address the concern that the routing of roads is endogenous. Our instrumental variable strategy is rooted in the observation that Roman roads were predominantly built for military purposes, i.e. to facilitate timely supply shipments and bringing in reinforcements (see Section 2).

F.1 Construction of instrument

To construct our instrument, we draw on the comprehensive dataset of Roman battles collected in Adamson (2020). In total, our dataset contains information on the location of 594 battles.[62,63] Based on these battle sites, which we use as nodes, we construct a network of artificial roads that connect the battle grounds to the capital Rome. Consequently, we add Rome to the set of nodes. Also defined as nodes are all harbours because military troops were frequently shipped via sea. The resulting set of nodes form the anchor points based on which we create our artificial road network.

One way of minimising travel times for troops between the various battle sites and Rome would be to created direct bilateral connections between any node pair. However, this would also maximise construction cost. To incorporate the trade-off between travel speed and construction costs and to avoid having to make arbitrary choices about which nodes to connect, we use the well-known Gabriel graph methodology (Gabriel and Sokal, 1969) to create the road network. This algorithm directly connects nodes in straight lines only if they are deemed to be 'neighbours'. Specifically, two nodes (n_1 and n_2) are joined by a straight line segment if and only if all other nodes lie outside the $n_1 - n_2$-circle (i.e., the circle on whose circumference n_1 and n_2 are at opposite points). The resulting road network along with the nodes is depicted in Panel (a) of Figure F.1, where the figure is restricted to the geographical scope of our analysis.

To create the instrumental variable, the Gabriel straight-line network is superimposed on the unmodified network of navigable rivers and coastal sea routes. In this combined network, transshipment between roads and other modes is allowed where the Gabriel road network intersects rivers or coasts. The final IV network is depicted in Panel (b) of Figure F.1. In analogy to the methodology described in Section 3, we use the Roman-technology-specific freight rates ($\boldsymbol{\alpha}^{Rome}$) to assign transport costs to the different network segments depending on their associated mode of transport. In the next step, we use Dijkstra (1959)'s algorithm to identify the least cost paths within our IV network. The costs associated with transport along the optimal paths are used as instrumental variable for the potentially endogenous Roman effective distance measure. Because the Gabriel straight-line network is less dense than the actual Roman road network, not all grid cells are intersected by the IV network. For grid cells that are not connected to the artificial network, we can therefore not predict Roman effective distance. As a consequence, the number of observations used in the IV approach is lower compared to our main analysis.

F.2 Estimation and results

To estimate the instrumental variable Poisson model, we specify it as an exponential conditional mean model with additive zero-mean error and with Roman effective distance allowed to be

[62] 304 of the 594 battle sites lie within Western Europe.

[63] Note that when coordinates for the location of a battle were missing in Adamson (2020), we added this information based on our own investigation.

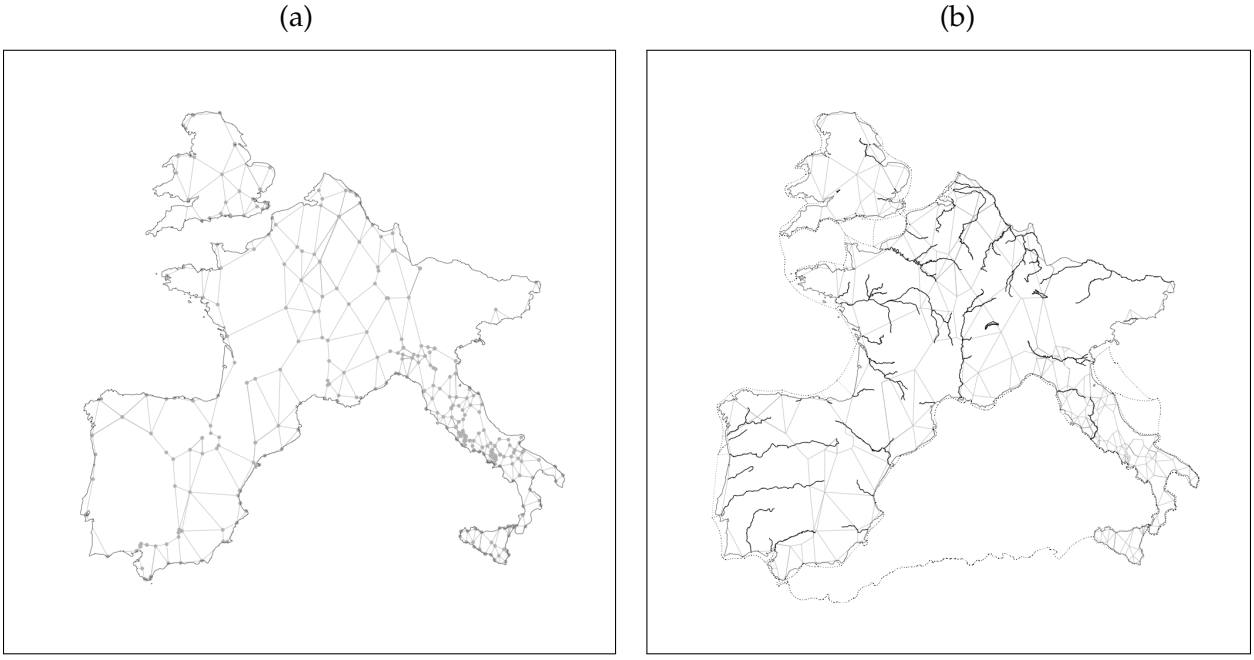


Figure F.1: *IV transport network*
Note. Panel (a) depicts the straight-line road network (grey lines) resulting from connecting the nodes (grey dots) using the Gabriel algorithm. Panel (b) shows the combined multi-modal IV transport network. Grey lines symbolise roads created by connecting the nodes (grey dots) using the Gabriel graph algorithm. Solid black lines represent navigable river sections, and dashed lines coastal shipping routes. Figures are restricted to the geographical scope of our analysis.

endogenous:

$$\mathcal{X}_{ij} = \exp\left(\delta \ln LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij} + \theta\, d_{ij} + T'_{ij}\,\gamma + \beta_i + \beta_j\right) + \varepsilon_{ij}. \tag{F.1}$$

The error function can then be written as:

$$\begin{aligned} &\varepsilon(\mathcal{X}_{ij}, \ln LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij}, d_{ij}, T_{ij}, \beta_i, \beta_j) \\ &= \mathcal{X}_{ij} - \exp\left(\delta \ln LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij} + \theta\, d_{ij} + T'_{ij}\,\gamma + \beta_i + \beta_j\right). \end{aligned} \tag{F.2}$$

We can then formulate the population moment conditions as:

$$E\left[\mathbf{z}\varepsilon(\mathcal{X}_{ij}, \ln LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij}, d_{ij}, T_{ij}, \beta_i, \beta_j)\right] = \mathbf{0}, \tag{F.3}$$

where the vector $\mathbf{z}$ is the vector of instruments that contains all exogenous variables and replaces the potential endogenous regressor $\ln LC(\mathbf{N}^{Rome}, \boldsymbol{\alpha}^{Rome})_{ij}$ with the instrument.

This model could in principle be estimated with pre-implemented commands, such as `ivpoisson` in Stata using two-step Generalised Method of Moments (GMM) estimation. However, note that we include origin and destination fixed effects which leads to many regressors and slows down estimation to the extent that it is practically infeasible. Using recent developments for linear (e.g. Gaure, 2013; Correia, 2016) and generalised linear models with high-dimensional fixed effects (e.g. Stammann, 2017; Bergé, 2018; Correia, Guimarães and Zylkin, 2020), feasibility of the estimation of the instrumental variable Poisson model can be obtained by first-order Taylor

approximation of the moment condition and reformulating such that the Method of Alternating Projections (MAP) can be used (see Hudlet and Larch, 2020). Basically, MAP can be thought of as a generalisation of the within-groups transformation, where the group-wise demeaning of variables is done in an alternating fashion, thereby avoiding the brute force use of dummies for the exporter and importer fixed effects (see Gaure, 2013).

Table F.1 presents the regressions of Roman effective distance on the IV as well as the and reduced-form estimates. Note that our IV estimates are obtained using GMM as we have a non-linear model and a two-step procedure would lead to inconsistent estimates (see Cameron and Trivedi, 2005). A first-stage does therefore not exist. The first two columns focus on the Roman era. Clearly, the instrument strongly predicts Roman effective distance (column 1) and deters trade in pottery (column 2). In columns (3)–(4) we repeat the exercise for the modern era; the pattern of results is the same.

Table F.1: Effect of IV on Roman effective distance and ownership intensity

	Rome		Current-Day	
Dependent Variable:	Roman effective distance	Share Pottery Finds	Roman effective distance	Share Ownership Links (>25% Ownership)
	'Relevance'	Reduced form	'Relevance'	Reduced form
	(1)	(2)		
Instrument	0.698***	- 1.454***	0.695***	-0.335***
	(0.028)	(0.275)	(0.010)	(0.069)
ln geodesic distance	0.044**	-0.861***	0.024***	-1.515***
	(0.017)	(0.300)	(0.004)	(0.065)
Baseline controls	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes
Observations	15,698	15,698	442,183	442,183
Estimator	PPML	PPML	PPML	PPML

Notes: This table reports estimates of the endogenous variable (Roman effective distance) on the instrumental variable in columns (1) and (3). Columns (2) and (4) report reduced-form estimates of the share of terra sigillata finds or the share of (partly) owned firms on the instrumental variable. The instrumental variable is a measure of effective distance that replaces roads with straight-line segments from a Gabriel graph, as described in detail in Appendix F.1. Baseline controls in columns (1–2) correspond to column 2 in Table 2 and in columns (3–4) correspond to column 2 in Table 3. They are described in the respective table notes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G Robustness tests: least-costs paths and geography

G.1 Construction of geography-based least-cost measures

This appendix describes the construction of geography-based least-cost paths and presents robustness tests that control for these measures. To show that our main estimates are not biased due to unobserved heterogeneity in geography, we account for a number of geography-based least-cost measures. Each measure is derived from a a 1km×1km gridded cost surface that captures how expensive it is to cross a cell. Given this surface, we determine the least-cost path between the centroids of any two 0.5×0.5 grid cells by applying Dijkstra (1959)'s algorithm. The associated cost is then defined as the geography-based connectivity measure. We construct two types of measures: measures that reflect energy expenditure and measures that capture travel time.

Energy expenditure We construct four separate metabolic energy expenditure cost surfaces. Three surfaces are derived from topography-based functions that predict how much energy humans expend when walking across different types and slopes of terrain. The specific functions we use were developed in Pandolf, Givoni and Goldman (1977), van Leusen (2002), and Llobera and Sluckin (2007).[64] The fourth energy expenditure cost surface is based on the wheeled-vehicle critical slope cost function introduced in Herzog (2013). This function was specifically tailored to wheeled vehicles (i.e., carts) by taking into account that these cannot climb steep slopes as easily as walkers can. The overland cost surfaces are constructed based on the R-package 'movecost' developed by Alberti (2019). All energy expenditure functions used in our analysis are listed in Alberti (2019).

Unfortunately, none of the four energy-expenditure functions model how costly it is to use waterborne transport methods. We thus have to make an assumption about the (relative) costs of moving on rivers and across sea. We take a very conservative stance by assuming that relative energy expenditure is equal to the mode-dependent transport cost ratio of the Roman era.[65] That is, coastal transport is 52 times cheaper than moving over a plain land surface; riverine transport is 7.5 times more expensive than coastal shipping.

Travel time We construct two travel-time based measures of connectivity. The first measure is the Human Mobility Index with Seafaring (HMISea) developed in Özak (2018).[66] The HMISea represents a measure of travel time in the pre-industrial era and takes human biological constraints as well as geographical and technological factors into account.[67] The second connectivity measure is based on the topography-influenced travel time function of Langmuir (2003, pp. 39 ff.). Following Barjamovic et al. (2019) we assume that travel over rivers, lakes, or sea is 10% faster than over a featureless plain overland.

G.2 Extended set of geographical and climatic controls

To account for the possibility that Roman connectivity partly reflects pre-existing geographical differences and obstacles between origins and destinations, we control for an extended set of

[64] The functions of Pandolf, Givoni and Goldman (1977) and van Leusen (2002) measure cost in Watts, whereas Llobera and Sluckin (2007) measures costs in Kilojoule per metre.

[65] In essence, we assume that transport technologies are the same as in the Roman era. More precisely, that the available technologies result in the same relative difference in shipping costs across modes.

[66] We are grateful to Ömer Özak for providing us with the data.

[67] See Özak (2010) and Özak (2018, p. 191), for a detailed description of the index.

controls that are described below.

Absolute difference in longitude Continuous variable that measures the absolute difference in decimal degrees between the longitude of the centroids of grid cells i and j.

Absolute difference in elevation Continuous variable that measures the absolute difference in elevation in meters between the centroids of grid cells i and j based on data from WorldClim (v. 2.1).

Absolute difference in ruggedness Continuous variable that measures the absolute difference in ruggedness within grid cells i and j. Ruggedness is based on data from WorldClim (v. 2.1) and calculated using the index devised in Riley, Degloria and Elliot (1999).[68]

Absolute difference in agricultural suitability Continuous variable that measures the absolute difference in the the Caloric Suitability Index (CSI) between grid cells i and j. The CSI was developed in Galor and Özak (2016) and captures the average potential yields within each cell attainable given the set of crops that are suitable for cultivation in the pre-1500 period.

Absolute difference in precipitation Continuous variable that measures the absolute difference in the average precipitation in millimetres between grid cells i and j over the 1970–2000 time horizon, obtained from the WorldClim (v. 2.1) data set.

Absolute difference in temperature Continuous variable that measures the absolute difference in the temperature in degrees Celsius between grid cells i and j over the 1970–2000 time horizon, obtained from the WorldClim (v. 2.1) data set.

Absolute difference in access to waterways Continuous variable that measures the absolute difference in the Strahler number between grid cells i and j, obtained from Vörösmarty et al. (2000). The Strahler number, also known as Strahler stream order, is a positive whole number used in geomorphology and hydrology to indicate the level of branching in a river system.

Average ruggedness on straight line Continuous variable that measures the average ruggedness (as defined in Riley, Degloria and Elliot, 1999) measured on a straight line drawn between centroids of grid cells i and j, derived from elevation data taken from WorldClim (v. 2.1).

River intersects straight line Indicator variable that assumes the value one if a straight line drawn between centroids of grid cells i and j intersects with a river included in the Roman transport network.

Coastline intersects straight line Indicator variable that assumes the value one if a straight line drawn between centroids of grid cells i and j intersects with a coastline.

Same watershed Indicator variable that assumes the value one if both grid cells i and j are located on the same watershed, obtained from the ECRINS (v. 1.1) data set.

[68] This index is, for example, also used in Nunn and Puga (2012).

Least-cost path runs via road Indicator variable that assumes the value one if the least-cost path within the Roman transport network (Roman effective distance) between grid cells i and j runs via a road segment.

Least-cost path runs via river Indicator variable that assumes the value one if the least-cost path within the Roman transport network (Roman effective distance) between grid cells i and j runs via a river segment.

Least-cost path runs via sea Indicator variable that assumes the value one if the least-cost path within the Roman transport network (Roman effective distance) between grid cells i and j runs via a coastal shipping route segment.

G.3 Results

Table G.1 documents the stability of our Roman-era result with respect to the inclusion of the geography controls introduced above. Compared to the baseline estimate (presented in column (4) of Table 2), the coefficient of Roman effective distance remains stable when we successively control for geography-based least-cost measures described above (columns 1–6). The point estimate also changes very little when we control for an extensive set of bilateral differences in geographical and climatic aspects (column 7). In column (8) we simultaneously account for all least-cost measures as well as the extended set of controls. Reassuringly, the point estimate of Roman effective distance also remains stable in this case.

In Table G.2, we investigate the robustness of our current-day estimates. In line with the picture above, we find that the Roman effective distance coefficient changes only slightly when we account for alternative least-cost measures and/or the extended set of geography controls.

Table G.1: Robustness geography–Roman era

Dependent Variable:	Share of Terra Sigillata Finds							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln Roman effective distance	-1.649***	-1.675***	-1.603***	-1.641***	-1.605***	-1.258**	-1.258***	-1.033***
	(0.631)	(0.636)	(0.621)	(0.634)	(0.502)	(0.529)	(0.330)	(0.348)
ln geodesic distance	-0.772**	-0.795**	-0.732**	-0.768**	0.899*	7.757***	-0.610**	3.030
	(0.374)	(0.383)	(0.361)	(0.374)	(0.503)	(1.573)	(0.250)	(1.877)
Least-cost measure	(a)	(b)	(c)	(d)	(e)	(f)	None	All
Extended set of geography controls	No	No	No	No	No	No	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,839	22,839	22,839	22,839	22,839	22,839	22,839	22,839
Estimator	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid-cell level are reported in parentheses. Dependent variable is the share of terra sigillata finds in cell i that originates from cell j. Baseline controls correspond to column 2 in Table 2 and are described in the respective table notes. Columns (1) to (6) alternate between controlling for various least cost measures described in detail in Appendix G. Column (1) uses measure (a) from Llobera and Sluckin (2007); Column (2) uses measure (b) from van Leusen (2002); Column (3) uses measure (c) from Pandolf, Givoni and Goldman (1977); Column (4) uses measure (d) from Herzog (2013); Column (5) uses measure (e) from Özak (2018); Column (6) uses measure (f) from Langmuir (2003). Column (7) adds the extended set of geography controls that includes the absolute difference in longitude, elevation, ruggedness, agricultural suitability, precipitation, temperature, and access to waterways (measured by the Strahler Index). Furthermore, it includes the average ruggedness on the straight line, two variables indicating whether the straight line intersects a river or a coastline, respectively. Finally, it includes a variable that indicates whether two grid cells are located in the same watershed and three variables that indicate whether the least cost path between two grid cells runs via any river, road, and sea segments, respectively. Column (8) adds all least cost measures from columns (1)–(6) and geography controls from column (7). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table G.2: Robustness geography–Today

Dependent Variable:	Share of Ownership Links (>25% Ownership)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln Roman effective distance	-0.526***	-0.526***	-0.527***	-0.524***	-0.435***	-0.434***	-0.410***	-0.503***
	(0.081)	(0.081)	(0.081)	(0.081)	(0.077)	(0.076)	(0.077)	(0.080)
ln geodesic distance	-1.620***	-1.621***	-1.617***	-1.619***	-1.433***	-1.673***	-1.613***	-1.665***
	(0.067)	(0.067)	(0.066)	(0.067)	(0.094)	(0.350)	(0.070)	(0.363)
Least-cost measure	(a)	(b)	(c)	(d)	(e)	(f)	None	All
Extended set of geography controls	No	No	No	No	No	No	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	731,823	731,823	731,823	731,823	731,823	731,823	731,823	731,823
Estimator	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table reports estimates of Equation (1) using the PPML estimator. Standard errors two-way clustered at the origin and destination grid-cell level are reported in parentheses. Dependent variable is the share of firms in cell j that are (partly) owned by firms located in cell i. Baseline controls correspond to column 2 in Table 3 and are described in the notes of Tables 2 and 3. Columns (1) to (6) alternate between controlling for various least cost measures described in detail in Appendix G. Column (1) uses measure (a) from Llobera and Sluckin (2007); Column (2) uses measure (b) from van Leusen (2002); Column (3) uses measure (c) from Pandolf, Givoni and Goldman (1977); Column (4) uses measure (d) from Herzog (2013); Column (5) uses measure (e) from Özak (2018); Column (6) uses measure (f) from Langmuir (2003). Column (7) adds the extended set of geography controls that includes the absolute difference in longitude, elevation, ruggedness, agricultural suitability, precipitation, temperature, and access to waterways (measured by the Strahler Index). Furthermore, it includes the average ruggedness on the straight line, two variables indicating whether the straight line intersects a river or a coastline, respectively. Finally, it includes a variable that indicates whether two grid cells are located in the same watershed and three variables that indicate whether the least cost path between two grid cells runs via any river, road, and sea segments, respectively. Column (8) adds all least cost measures from columns (1)–(6) and geography controls from column (7). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

H Falsification test: network construction and summary statistics

This appendix provides a detailed description of how we construct the transport networks inside and outside of Western Europe for the falsification exercise. The falsification exercise is rooted in the idea that Roman transport costs should matter for today's integration only in regions that historically were part of the Roman territory (and therefore the Roman transport network).

In the absence of information on historical road networks for areas outside of the Roman territory, we draw on today's primary roads (highways). This is motivated by the quantitative and qualitative evidence that current highways follow historical infrastructure, both in regions within and beyond the Roman Empire (e.g. Garcia-López, Holl and Viladecans-Marsal, 2015; Percoco, 2015; Redding and Turner, 2015; Hitchner, 2012). The two transport networks, one for Western Europe and one for non-Roman Europe, are constructed by combining information on location of large navigable rivers from WISE (2020) and major roads from ESRI (2020).[69] Shipment by sea is possible along the coast. Depending on the mode of transportation, individual segments of the network are assigned with the specific transport costs $\boldsymbol{\alpha}^{Rome} = (1, 7.5, 52)$. We identify the least-cost path between any two grid cells within the two respective territories using Dijkstra (1959)'s algorithm. Figure H.1 depicts the transport networks for both, regions once integrated and those never part of the Roman Empire.

Figure H.1: The Multi-Modal Transport Network

Map shows the transport network (restricted to the geographical scope of our analysis). Grey lines symbolise highways (ESRI, 2020), solid black lines navigable rivers (WISE, 2020), and dashed lines coastal shipping routes.

[69] The data can be downloaded at `http://bit.do/WiseRivers` (rivers) and `http://bit.do/worldroads` (roads).

The scope of our falsification test is restricted to European countries that were never part of Rome (except for Finland, Norway, and Sweden). In addition to Western Europe (i.e., the geographical scope of our main analysis), we thus include ownership information from Bureau van Dijk's Orbis database for the following countries: Austria, Belarus, Czech Republic, Denmark, Estonia, Germany, Hungary, Latvia, Lithuania, Moldova, the Netherlands, Poland, Romania, Slovakia, and the Ukraine. For regions within these countries that were never part of the Rome, we construct the measure of grid-cell pair business link intensity in analogy to the procedure described in Section 3. We include relevant ownership links between grid cells located within Western Europe and within non-Roman Europe but not across the two territories. Table H.1 reports summary statistics for the two datasets.

Table H.1: Descriptive statistics falsification datasets

Variable	Mean	Std. dev.	Min.	Max.	Obs.
	Europe once part of the Roman Empire				
Share ownership links	0.001	0.015	0	1	723,323
ln Roman effective distance	9.271	0.430	6.248	10.447	723,323
ln geodesic distance	6.639	0.683	3.061	7.803	723,323
Absolute distance latitude	5.102	3.666	0	18.457	723,323
Same biome	0.290	0.454	0	1	723,323
Access to a waterway	0.203	0.402	0	1	723,323
Both Mediterranean Sea	0.016	0.125	0	1	723,323
	Europe never part of the Roman Empire				
Share ownership links	0.001	0.020	0	1	414,451
ln Roman effective distance	10.110	0.635	5.655	11.397	414,451
ln geodesic distance	6.538	0.698	2.762	7.805	414,451
Absolute distance latitude	3.604	2.733	0	15.670	414,451
Same biome	0.366	0.482	0	1	414,451
Access to a waterway	0.144	0.351	0	1	414,451
Both Mediterranean Sea	0	0	0	0	414,451

Notes: This table presents descriptive statistics of variables used for the falsification exercise in Table 6.

I Mechanisms: data and further results

This appendix describes the construction of measures capturing our various mechanisms introduced in Section 7 and provides auxiliary results. Table I.1 presents the summary statistics for all variables used in Section 7 and in this Appendix.

I.1 Data construction for channels

Google driving distance Measures the length of the time-minimising driving route between grid cell centroids in kilometres, extracted from Google Maps. This metric captures variation arising from distance in the road network and differences in the speed of transport associated with different technologies (i.e. motorways, rural roads, etc.). The measure aims at capturing the cost of transporting goods and people using today's road network.

Rome2Rio travel time Measures the duration of the time-minimising route between grid cell centroids in hours, extracted from `rome2rio.com`. This metric captures variation arising from distance and available transport mode in the multi-modal passenger-transport network, in which passengers are allowed to use any combination of public transport (bus, train, aeroplane).[70]

Optimal Transport Routes Over Time Figure I.1 provides a visual comparison of optimal routes between Toulouse and Reims during Roman times (panel a), today's optimal routes via roads (panel b), and via public transport (panel c). The figure aims to support our interpretation that the correlation between historical and contemporary connectivity does not reflect persistence in the layout of the road network but rather that regions with historically stronger ties were connected more directly when new transport technologies became available (e.g., railways, aeroplanes, and highways). Thus, even though past and present transport networks structurally differ in their layout and transport technologies, Roman-era-specific connectivity still explains patterns of bilateral accessibility today.

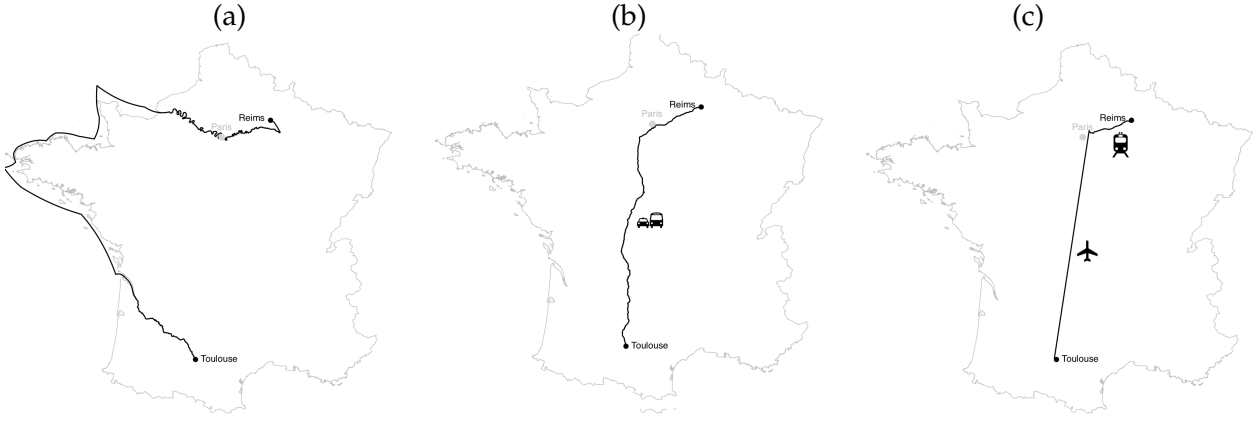


Figure I.1: Optimal Transport Routes Over Time
Panel (a) depicts the least-cost path between Toulouse and Reims within the Roman transport network; panel (b) shows time-minimising route using today's road network (Google Maps); panel (c) shows time-minimising path within today's passenger transport network (`www.rome2rio.com`).

[70] We also allow for taxi rides when public transport is not available.

Industry dissimilarity We draw on Bureau van Dijk's Orbis database to construct a grid-cell-pair-level measure of industry dissimilarity. The procedure for constructing the measure is analogous to locating firms in patent-technology space developed in Jaffe (1986) in order to quantify R&D spillovers. In a first step we aggregate firms in a given grid cell to the three-digit NACE industry classification level, i.e., we create a vector of grid-cell-by-industry counts of firms. Next, we calculate a vector of the share of firms $F_i = (F_{i1}, F_{i2}, ..., F_{i344})$ in grid-cell i that are operating in industry x. In a final step, we produce a matrix of bilateral industry dissimilarity S, defined as the uncentred correlation of industry share vectors between any two grid-cells i and j, using the methodology established by Jaffe (1986):

$$S_{ij} = 1 - \frac{F_i F_j'}{(F_i F_i')^{\frac{1}{2}} (F_j F_j')^{\frac{1}{2}}}.$$

This measure lies between zero and one. It takes a value of zero for grids-cell pairs whose vectors of industry shares is identical and one for grid-cell pairs whose vectors of industry pairs is orthogonal, i.e., higher values reflect greater dissimilarity in the industry structure.[71] Interestingly, the measure has the advantage that it is not affected by the length of the F vectors, i.e., it will be robust to aggregation of classification levels such as moving from NACE 3 to NACE 2, whereas other distance measures might be sensitive to the distance between vector endpoints.

Preferences, attitudes and values To construct grid-cell-pair-level measures of preference (dis-) similarity, we draw on the Global Preferences Survey (GPS, Falk et al., 2018). The GPS is an experimentally validated dataset—described in detail in Appendix A of Falk et al. (2018)—that collects 6 measures specifically designed to capture time preferences, risk preferences, positive and negative reciprocity, altruism, and trust. The survey covers all Western European countries except for Belgium and Luxembourg.

To construct grid-cell-pair-level measures of attitude and value (dis-)similarity, we draw on the European Values Study (EVS wave 4, 2008). The EVS is a research survey program that covers all Western European countries and elicits information on ideas, beliefs, preferences, attitudes, values, and opinions of citizens in all Western European countries using a battery of questions. Following the EVS categorization, we group the questions into six topics, each covering a specific aspect of values and attitudes. The topics, along with the associated questions, can be found at `http://bit.do/evstopics`. For manageability reasons, we run a multiple correspondence analysis and compute the row coordinates of the first dimension within each topic and use the resulting values as the basis for our grid-cell-pair-level measures of similarity.

The procedure for constructing grid-cell-pair similarity measures from the individual-level information is the same for the GPS and the EVS surveys. In a first step, we purge all variables of country fixed effects. We then identify in which NUTS2 region a respondent currently lives (GPS) or where she/he resided at age 14 (EVS). NUTS2 regions are the most detailed spatial information on residence available in the surveys. In a third step, we compute the average values of the preference and value measures for each NUTS2 region. These values are then assigned to the grid cells. If a grid cell falls entirely within one single NUTS2 region, it is simply assigned the NUTS2 level average of the preference or value measure. If a grid cell overlaps multiple NUTS2 regions, we determine the share of total grid cell population that lives within each of

[71] For ease of interpretation, we have adapted the inverse Jaffe (1986) index to reflect dissimilarity. The original Jaffe (1986) index (O_{ij}), captures similarity and is given by: $O_{ij} = \frac{F_i F_j'}{(F_i F_i')^{\frac{1}{2}} (F_j F_j')^{\frac{1}{2}}}$.

the intersecting NUTS2 region and compute the population weighted averages of the respective preference and value measures. Formally, the population weighted measure M_i for a given grid cell i is defined as:

$$M_i = \sum_{n \in N_i} p_n m_n.$$

The proportion of total grid cell population living in NUTS2 region n is represented by p_n, and m_n captures the NUTS2-level average of a given preference or value measure. Population data are taken from Schiavina, Freire and MacManus (2019).

In the final step, we compute the grid-cell-pair-level distance indices as the absolute difference between two grid cells in the respective preference, value or attitude measure.

Social connectedness Our measure for the bilateral intensity of social ties is the Social Connectedness Index (SCI) developed in Bailey et al. (2018). The SCI captures the link strength between two regions within the Facebook social network.[72] For Western Europe, the SCI is available at the NUTS3-pair level. When both grid cells fall entirely within one NUTS3 region, we use the corresponding NUTS3-pair level SCI as measure for social connectedness. If (one of) the cells overlaps multiple NUTS3 regions, we compute the population-weighted SCI.[73] This population weighted SCI between grid cell i and j is defined as:

$$SCI_{i,j} = \sum_{r \in N_i} \sum_{q \in N_j} p_r p_q sci_{r,q}$$

where the proportion of grid cell i's total population that lives in NUTS3 region r is represented by p_r and proportion of grid cell j's total population that lives in NUTS3 region q is represented by p_q; $sci_{r,q}$ is the SCI for the NUTS3 pair r and q. Using unweighted averages or assigning grid-cell pairs to the SCI based on their centroids produces very similar results.

I.2 Results for preference and value mechanisms by group

In Table I.2, we analyse the effects of Roman network connectivity separately for the individual preferences and value groups. We find that dissimilarity in the majority of aspects increases with Roman effective distance. This highlights the importance of considering socio-economic forces, in our case the transport-network-induced cumulative history of exchange between regions, when trying to understand why preferences and attitudes vary across regions.

In Table I.3, we add all individual measures of preferences and values and attitudes instead of aggregate measures. The results are qualitatively similar to the main results in Table 8.

[72] Note that the data released by Facebook are scaled by an (unknown) factor. Therefore absolute values of the SCI are not meaningful in the sense that they cannot be directly be interpreted as the number of Facebook user links between regions.

[73] Population data are taken from Schiavina, Freire and MacManus (2019).

Table I.1: Descriptive statistics of key variables and potential mechanisms

Variable	Mean	Std. dev.	Min.	Max.	Obs.
			Main variables		
Share ownership links	0.001	0.016	0	1	674,805
ln Roman effective distance	9.404	0.474	5.953	10.658	674,805
ln geodesic distance	6.648	0.684	3.061	7.802	674,805
Intra-national ownership	0.194	0.395	0	1	674,805
Same biome	0.299	0.458	0	1	674,805
Absolute distance latitude	5.109	3.669	0	18.457	674,805
Access to a waterway	0.200	0.400	0	1	674,805
Both Mediterranean Sea	0.017	0.131	0	1	674,805
Joint duration under Roman rule	4.413	0.705	3.570	6.720	674,805
			Potential mechanisms (as presented in Tables 7–8)		
ln Google driving distance (SD)	0	1	-5.249	1.814	674,805
ln Rio2Rome travel time (SD)	0	1	-8.903	5.426	674,805
Industry dissimilarity (SD)	0	1	-4.979	1.241	674,805
Principal component preferences (SD)	0	1	-2.380	6.750	674,805
Principal component attitudes (SD)	0	1	-2.191	5.997	674,805
ln SCI (SD)	0	1	-1.763	6.097	674,805
			Potential mechanisms (as presented in Tables I.2–I.3)		
Trust (SD)	0	1	-1.175	6.431	674,805
Altruism (SD)	0	1	-1.188	8.104	674,805
Negative reciprocity (SD)	0	1	-1.273	8.723	674,805
Positive reciprocity (SD)	0	1	-1.204	6.262	674,805
Risk (SD)	0	1	-1.246	7.519	674,805
Time (SD)	0	1	-1.220	4.635	674,805
Life (SD)	0	1	-1.247	6.679	674,805
Work (SD)	0	1	-1.187	6.111	674,805
Family (SD)	0	1	-1.236	6.086	674,805
Politics and society (SD)	0	1	-1.129	7.823	674,805
Religion (SD)	0	1	-1.243	7.509	674,805
Nationalism (SD)	0	1	-1.210	6.124	674,805

Notes: This table presents descriptive statistics of variables used for the analysis in Tables 7–8 and Tables I.2–I.3. Mechanism variables are standardized with mean zero and standard deviation of one.

Table I.2: Roman transport network connectivity and current preferences and values by topic

Panel a: Distance in preferences

	Trust (SD)	Altruism (SD)	Negative reciprocity (SD)	Positive reciprocity (SD)	Risk (SD)	Time (SD)
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	0.031*	0.143***	0.072***	0.042***	0.037*	-0.012
	(0.018)	(0.020)	(0.021)	(0.013)	(0.020)	(0.018)
ln geodesic distance	0.185***	0.073***	0.045***	0.074***	0.208***	0.158***
	(0.017)	(0.012)	(0.009)	(0.014)	(0.017)	(0.013)
Raw mean of dep. var.	0.170	0.151	0.155	0.149	0.153	0.154
SD of raw dep. var.	0.145	0.127	0.122	0.124	0.123	0.126
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	674,805	674,805	674,805	674,805	674,805	674,805
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Panel b: Distance in values and attitudes

	Attitudes and Values towards:					
	Life (SD)	Work (SD)	Religion (SD)	Family (SD)	Politics and society (SD)	Nationalism (SD)
	(1)	(2)	(3)	(4)	(5)	(6)
ln Roman effective distance	-0.037*	0.135***	0.039***	0.100***	0.015	0.269***
	(0.022)	(0.027)	(0.013)	(0.026)	(0.017)	(0.049)
ln geodesic distance	0.145***	0.088***	0.081***	0.105***	0.109***	0.038**
	(0.016)	(0.010)	(0.010)	(0.015)	(0.014)	(0.015)
Raw mean of dep. var.	0.229	0.254	0.292	0.278	0.254	0.354
SD of raw dep. var.	0.183	0.214	0.235	0.225	0.225	0.293
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	674,805	674,805	674,805	674,805	674,805	674,805
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Notes: This table reports estimates of Equation (1) using the OLS estimator. Standard errors two-way clustered at the origin and destination grid-cell level are reported in parentheses. Panel (a) uses the six individual measures of preferences in the GPS as dependent variables. Panel (b) uses the first dimension of a multiple correspondence analysis executed within each of the topics of the EVS as dependent variables. For more details on the dependent variables, see Appendix I. Baseline controls correspond to column 2 in Table 3 and are described in the notes of Tables 2 and 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table I.3: Accounting for potential channels: preferences and values by topic

Dependent Variable:	Number of Ownership Links (>25% Ownership)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln Roman effective distance	-0.370***	-0.362***	-0.237***	-0.236***	-0.362***	-0.318***	-0.314***	-0.292***	-0.178**	-0.065
	(0.079)	(0.079)	(0.073)	(0.073)	(0.079)	(0.079)	(0.077)	(0.078)	(0.071)	(0.063)
ln Driving distance (SD)		-0.218*		-0.028					0.061	
		(0.127)		(0.130)					(0.124)	
ln Rome2Rio (SD)			-0.249***	-0.248***					-0.219***	
			(0.021)	(0.022)					(0.021)	
Industry dissimilarity (SD)					-0.122***				-0.114***	
					(0.024)				(0.023)	
Distance trust (SD)						-0.035		-0.007	-0.009	
						(0.033)		(0.032)	(0.032)	
Distance altruism (SD)						-0.185***		-0.158***	-0.138***	
						(0.023)		(0.022)	(0.021)	
Distance negative reciprocity (SD)						-0.058***		-0.024	-0.020	
						(0.022)		(0.021)	(0.020)	
Distance positive reciprocity (SD)						-0.083***		-0.066**	-0.048*	
						(0.026)		(0.027)	(0.027)	
Distance risk (SD)						-0.015		0.013	0.014	
						(0.029)		(0.029)	(0.030)	
Distance patience (SD)						-0.018		0.003	0.005	
						(0.025)		(0.025)	(0.024)	
Distance values & attitudes life (SD)							-0.159***	-0.146***	-0.146***	
							(0.029)	(0.029)	(0.028)	
Distance values & attitudes work (SD)							-0.107***	-0.087***	-0.085***	
							(0.029)	(0.029)	(0.028)	
Distance values & attitudes family (SD)							-0.066**	-0.050*	-0.047*	
							(0.028)	(0.027)	(0.027)	
Distance values & attitudes politics (SD)							-0.039	-0.014	-0.001	
							(0.028)	(0.027)	(0.027)	
Distance values & attitudes religion (SD)							-0.089***	-0.067***	-0.069***	
							(0.024)	(0.023)	(0.022)	
Distance values & attitudes nationalism (SD)							0.037	0.061*	0.068**	
							(0.035)	(0.036)	(0.033)	
ln SCI (SD)										1.140***
										(0.043)
ln geodesic distance	-1.541***	-1.243***	-1.160***	-1.122***	-1.534***	-1.408***	-1.400***	-1.366***	-1.131***	-0.344***
	(0.062)	(0.168)	(0.072)	(0.158)	(0.062)	(0.065)	(0.063)	(0.064)	(0.152)	(0.063)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	674,805	674,805	674,805	674,805	674,805	674,805	674,805	674,805	674,805	674,805
Estimator	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML	PPML

Notes: This table replicates Table 8 using individual preferences and value topics instead of aggregate measures using the PPML estimator. Standard errors two-way clustered at the origin and destination grid cell level are reported in parentheses. Dependent variable is the share of firms in cell j that are (partly) owned by firms located in grid i. Each column adds explanatory variables that serve as mechanisms to explain the results in Table 3. For details on these explanatory variables, see the main text and Appendix I. Baseline controls correspond to column 2 in Table 3 and are described in the notes of Tables 2 and 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Adamson, Jordan. 2020. "Political Institutions, Resources, and War: Theory and Evidence from Ancient Rome." *Explorations in Economic History* 76: 101324.

Alberti, Gianmarco. 2019. "movecost: An R Package for Calculating Accumulated Slope-dependent Anisotropic Cost-surfaces and Least-cost Paths." *SoftwareX* 10: 100331.

Archäologisches Landesmuseum Baden-Württemberg. 2006. *Imperium Romanum – Roms Provinzen an Neckar, Rhein und Donau: Begleitbuch zu Landesaustellung im Kunstgebäude Stuttgart vom 01.10.2005 - 08.01.2006.* Archäologisches Landesmuseum Baden-Württemberg.

Artioli, Gilberto, Caterina Canovaro, Paolo Nimis, and Ivana Angelini. 2020. "LIA of Prehistoric Metals in the Central Mediterranean Area: A Review." *Archaeometry* 62 (S1): 53–85.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives* 32 (3): 259–80.

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

Beck, Curt W., Constance A. Fellows, and Audrey B. Adams. 1970. "Analysis and Provenience of Minoan and Mycenaean Amber, III. Kakovatos." *Greek, Roman, and Byzantine Studies* 11 (1): 5–22.

Beck, Curt W., Gretchen C. Southard, and Audrey B. Adams. 1972. "Analysis and Provenience of Minoan and Mycenaean Amber, IV. Mycenae." *Greek, Roman, and Byzantine Studies* 13 (4): 359–385.

Begemann, Friedrich, Sigrid Schmitt-Strecker, Ernst Pernicka, and Fulvia Lo Schiavo. 2001. "Chemical Composition and Lead Isotopy of Copper and Bronze from Nuragic Sardinia." *European Journal of Archaeology* 4 (1): 43–85.

Benedictow, Ole Jørgen. 2006. *The Black Death, 1346-1353: The Complete History.* Boydell Press.

Bergé, Laurent. 2018. "Efficient Estimation of Maximum Likelihood Models with Multiple Fixed-Effects: The R Package FENmlm." *CREA Discussion Paper Series available for download at https://EconPapers.repec.org/RePEc:luc:wpaper:18-13*.

Biraben, Jean-Noël. 1975. *Les Hommes et la Peste en France et dans les Pays Europeens et Mediteraneens.* Mouton.

Boerner, Lars, and Battista Severgnini. 2014. "Epidemic Trade." London School of Economics Economic History Working Papers 212.

Burgess, Colin, and Brendan O'Connor. 2008. "Iberia, the Atlantic Bronze Age and the Mediterranean." *Contacto Cultural entre el Mediterráneo y el Atlántico (siglos XII-VIII Ane). La Precolonización a Debate. Serie Arqueológica* 11: 41–58.

Burriel Alberich, Josep, Albert Ribera i Lacomba, and María Serrano Marco. 2004. "A Fluvial Harbour of the Roman Period at Valentia (Hispania Tarraconensis)." In *Close Encounters: Sea and Riverborne Trade, Ports and Hinterlands, Ship Construction and Navigation in Antiquity, the Middle Ages and in Modern Time*, edited by Marinella Pasquinucci and Timm Weski, 129–138. Oxford University Press.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics - Methods and Applications.* Cambridge, United Kingdom: Cambridge University Press.

Campbell, Brian. 2012. *Rivers and the Power of Ancient Rome.* UNC Press Books.

Canovaro, Caterina, Ivana Angelini, Gilberto Artioli, Paolo Nimis, and Elisabetta Borgna.

2019. "Metal Flow in the Late Bronze Age across the Friuli-Venezia Giulia Plain (Italy): New Insights on Cervignano and Muscoli Hoards by Chemical and Isotopic Investigations." *Archaeological and Anthropological Sciences* 11 (9): 4829–4846.

Carreras, Cesar, and Pau De Soto. 2013. "The Roman Transport Network: A Precedent for the Integration of the European Mobility." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 46 (3): 117–133.

Carreras Monfort, César, and Pau De Soto. 2008/2009*b*. "La red comunicaciones romana en Cataluña: actualizacion y metodologia." *ANAS. Aspectos de la red viaria hispano-romana* 21–22: 319–340.

Carreras Monfort, César, and Pau De Soto. 2009*a*. "La Movilidad en Época Romana en Hispania: Aplicaciones de Análisis de Redes (SIG) Para el Estudio Diacrónico de las Infraestructuras de Transporte." *Habis* 40: 303–324.

Childe, Gordon. 1930. *The Bronze Age.* Cambridge University Press.

Childe, Gordon. 1958. *The Prehistory of European Society.* Penguin Books.

Christakos, George, Ricardo A. Olea, Marc L. Serre, Lin-Lin Wang, and Hwa-Lung Yu. 2005. *Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death.* Springer.

Cipolla, Carlo M. 1974. "The Plague and the Pre-Malthus Malthusians." *Journal of European Economic History* 3 (2): 277–284.

Correia, Sergio. 2016. "A Feasible Estimator for Linear Models with Multi-Way Fixed Effects." *unpublished manuscript available at http://scorreia.com/research/hdfe.pdf*.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson Estimation with High-Dimensional Fixed Effects." *The Stata Journal* 20 (1): 95–115.

Cravino, Javier, and Andrei A. Levchenko. 2017. "Multinational Firms and International Business Cycle Transmission." *The Quarterly Journal of Economics* 132 (2): 921–962.

Cumberlidge, Jane. 2009. *Inland Waterways of Great Britain.* London: Imray.

Cummings, Vicki, Magdalena S. Midgley, and Chris Scarre. 2015. "The Oxford Handbook of Neolithic Europe." edited by Chris Fowler, Jan Harding and Daniela Hofmann, Chapter Chambered Tombs and Passage Graves of Western and Northern Europe. Oxford University Press.

Dannell, Geoffrey, and Allard W. Mees. 2015. "Getting Samian Ware to Britain: Routes and Transport Possibilities." *Journal of Roman Pottery Studies* 16: 77–92.

Darvill, Timothy. 2013. "The Oxford Handbook of the European Bronze Age." edited by Harry Fokkens and Anthony Harding, Chapter Monuments and Monumentality in Bronze Age Europe. Oxford University Press.

De Soto, Pau. 2013*a*. "El Sistema de Transportes del Suroeste Peninsular en Época Romana. Análisis de del Funcionamiento de sus Infraestructuras." *Actas del Congreso Internacional VI Encuentro de Arqueologí del Suroeste peninsular* 1551–1576.

De Soto, Pau. 2013*b*. "O irado Mar Atlântico: o naufragio Bético Augustano de Esposende (Norte de Portugal)." *O Irado Mar Atlantico. O Naufragio Betico Augustano de Esposende (Norte de Portugal)*, edited by Rui Morais, Helena Granja and Angel Morillo, Chapter Los sistemas de transporte romanos y la configuración territorial en el noroeste peninsular, 179–210. Braga.

Dijkstra, Edsger W. 1959. "A Note on Two Problems in Connexion with Graphs." *Numerische Mathematik* 1 (1): 269–271.

Duncan-Jones, Richard. 1982. *Economy of the Roman Empire.* Cambridge University Press.

Eaton, Jonathan, Samuel S. Kortum, and Sebastian Sotelo. 2013. "International Trade: Linking

Micro and Macro." In *Advances in Economics and Econometrics: Tenth World Congress*, Vol. 3 of *Econometric Society Monographs*, edited by Daron Acemoglu, Manuel Arellano and Eddie Dekel. Cambridge University Press.

Eckoldt, Martin. 1983. "Schiffahrt auf kleinen Flüssen. Der Neckar und seine Nebenflüsse zur Römerzeit." *Deutsches Staatsarchiv* 6: 11–24.

Eckoldt, Martin. 1986. "Schiffahrt auf kleinen Flüssen. Nebenflüsse des Oberrheins und des unteren Mains im ersten Jahrtausend n. Chr." *Deutsches Staatsarchiv* 9: 59–88.

ESRI. 2020. "World Roads." http://bit.do/worldroads.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–1692.

Federico, Giovanni, Max-Stephen Schulze, and Oliver Volckart. forthcoming. "European Goods Market Integration in the Long Run." *Journal of Economic History*.

Gabriel, K. Ruben, and Robert R. Sokal. 1969. "A New Statistical Approach to Geographic Variation Analysis." *Systematic Biology* 18 (3): 259–278.

Gale, Noel H. 2006. "Lead Isotope Studies-Sardinia and the Mediterranean: Provenance Studies of Artefacts Found in Sardinia." *Instrumentum* 23: 29–34.

Gale, Noel H., and Zofia A. Stos-Gale. 1981. "Lead and Silver in the Ancient Aegean." *Scientific American* 244 (6): 176–193.

Gale, Noel H., and Zofia A. Stos-Gale. 2005. "Zur Herkunft der Kupferbarren aus dem Schiffwrack von Uluburun und der Spätbronzezeitliche Metallhandel im Mittelmeerraum." 117–132. Deutsches Bergbau-Museum.

Gale, Noel H., Zofia A. Stos-Gale, and Jack L. Davis. 1984. "The Provenance of Lead Used at Ayia Irini, Keos." *Hesperia: The Journal of the American School of Classical Studies at Athens* 53 (4): 389–406.

Galor, Oded, and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review* 106 (10): 3064–3103.

Garcia, João C. 1982. "Navegabilidade e Navegação no Baixo Guadiana." Centro de estudos geográficos phdthesis, Lisbon.

Garcia-López, Miquel-Àngel, Adelheid Holl, and Elisabet Viladecans-Marsal. 2015. "Suburbanization and Highways in Spain When the Romans and the Bourbons Still Shape Its Cities." *Journal of Urban Economics* 85: 52–67.

García y Bellido, Antonio. 1944. "La Navegación Ibérica en la Antigüedad, Según los Textos Cisicos y la Arqueologia." *Estudios Geográficos* 5 (16): 511–560.

Gattermayr, W., and D. Steck. 2006. "Innsbruck und das Hochwasser. Geschichte und Geschichten geschrieben vom Inn." Abt. Wasserwirtschaft beim Amt der Tiroler Landesregierung.

Gaure, Simen. 2013. "OLS with Multiple High Dimensional Category Variables." *Computational Statistics & Data Analysis* 66: 8–18.

Giumlia-Mair, Alessandra. 2005. "Handel und Rohstoffgewinnung im Italien der Späten Bronzezeit." 415–430. Deutsches Bergbau-Museum.

Hartley, Brian R., Brenda M. Dickinson, Geoffrey B. Dannell, Michael Fulford, Allard W. Mees, Paul Tyers, and Rosemary Wilkinson. 2008. *Names on Terra Sigillata. an Index of Makers' Stamps & Signatures on Gallo-roman Terra Sigillata (Samian Ware).* Vol. 1-9. University of London, Institute of Classical Studies.

Hauptmann, Andreas, Friedrich Begemann, and Sigrid Schmitt-Strecker. 1999. "Copper Objects from Arad: Their Composition and Provenance." *Bulletin of the American Schools of Oriental Research* 314 (1): 1–17.

Herzog, Irmela. 2013. "The Potential and Limits of Optimal Path Analysis." In *Computational Approaches to Archaeological Spaces*, edited by Andrew Bevan and Mark Lake, 179–211. Routledge.

Hitchner, R. Bruce. 2012. "Roads, Integration, Connectivity, and Economic Performance in the Roman Empire." In *Highways, Byways, and Road Systems in the Pre-Modern World*, edited by Richard J. A. Talbert Susan E. Alcock, John Bodel, 222–234. Wiley Blackwell.

Holst, Mads Kähler. 2013. "The Oxford Handbook of the European Bronze Age." edited by Harry Fokkens and Anthony Harding, Chapter Burials. Oxford University Press.

Hopkins, Keith. 1983. "Models, Ships and Staples." In *Trade and Famine in Classical Antiquity, Proceedings of the Cambridge Philological Society*, edited by P. D. A. Garnsey and C. R. Whittaker, 84–109. Cambridge: Philological Society Cambridge.

Hudlet, Alexander, and Mario Larch. 2020. "Fast Instrumental Poisson Estimation with High-Dimensional Fixed Effects." unpublished manuscript.

Jaffe, Adam. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review* 76 (5): 984–1001.

Kaiser, Alaina. 2013. "Copper Oxhide Ingot Marks: A Database and Comparative Analysis." Master's diss. Cornell University.

Klassen, Lutz, Serge Cassen, and Pierre Pétrequin. 2012. "Alpine Axes and Early Metallurgy." 1280–1309. Centre de Recherche Archéologique de la Vallée de l'Ain.

Kristiansen, Kristian, and Paulina Suchowska-Ducke. 2015. "Connected Histories: The Dynamics of Bronze Age Interaction and Trade 1500–1100 BC." *Proceedings of the Prehistoric Society* 81: 361–392.

Langmuir, Eric. 2003. *Mountaincraft and Leadership.* Vol. 3 edition. Manchester & Aviemore: British Mountaineering Council.

Latour, Louis. 2006. "Les fouilles gallo-romaines d'Auterive (Haute-Garonne), étude des couches les plus récentes." In *Mémoires de la Société Archéologique du Midi de La France*, 41–71.

Ling, Johan, Eva Hjärthner-Holdar, Lena Grandin, Zofia A. Stos-Gale, Kristian Kristiansen, Anne L. Melheim, Gilberto Artioli, Ivana Angelini, Rüdiger Krause, and Caterina Canovaro. 2019. "Moving Metals IV: Swords, Metal Sources and Trade Networks in Bronze Age Europe." *Journal of Archaeological Science: Reports* 26: 1–34.

Ling, Johan, Zofia A. Stos-Gale, Lena Grandin, Kjell Billström, Eva Hjärthner-Holdar, and Per-Olof Persson. 2014. "Moving Metals II: Provenancing Scandinavian Bronze Age Artefacts by Lead Isotope and Elemental Analyses." *Journal of Archaeological Science* 41: 106–132.

Llobera, M., and T.J. Sluckin. 2007. "Zigzagging: Theoretical insights on climbing strategies." *Journal of Theoretical Biology* 249 (2): 206 – 217.

Lo Schiavo, Fulvia. 2005. "Metallhandel im Zentralen Mittelmeer." 399–414. Deutsches Bergbau-Museum.

Mason, Andrea H. 2020. "Provenance of Tin in the Late Bronze Age Balkans: Preparation of Cassiterite for Sn Isotope Analysis and the Probabilistic and Spatial Analysis of Sn Isotopes." PhD diss. City University of New York.

Matthäus, Hartmut. 2005. "Kulturaustausch, Handel und Seefahrt im Mittelmeerraum während der Späten Bronzezeit." 333–366. Deutsches Bergbau-Museum.

Melheim, Anne L., Lena Grandin, Per-Olof Persson, Kjell Billström, Zofia A. Stos-Gale, Jo-

han Ling, Alan Williams, Ivana Angelini, Caterina Canovaro, Eva Hjärthner-Holdar, and Kristian Kristiansen. 2018. "Moving Metals III: Possible Origins for Copper in Bronze Age Denmark Based on Lead Isotopes and Geochemistry." *Journal of Archaeological Science* 96: 85–105.

Montero Ruiz, Ignacio, Isabel Martínez Navarrete, and Eduardo Galán. 2016. "Objetos o Materia Prima: Problemas en la Interpretación de Procedencias con Análisis de Isotopos de Plomo." *Boletín del Museo Arqueológico Nacional* 34: 81–98.

Montero Ruiz, Ignacio, María R. Manunza, Fulvia Lo Schiavo, Paolo Valera, José I. Ibarguchi, Nuria Rafel Fontanals, and Pau Sureda. 2018. "The Funtana Coberta-Ballao Hoard: New Copper Provenances in Nuragic Metallurgy." Vol. 56 of *Monographies Instrumentum*, 137–164. Éditions Mergoil.

Moret, Pierre. 2015. "Strabon et les Fleuves Gaulois." In *Les Gaulois au fil de leau. Actes du 37e Colloque international de lAFEAF (Montpellier, 8-11 mai 2013)*, edited by Fabienne Olmer and Rejane Roure, 217–234. ausonius.

Munoz, Alicia Fornell. 1997. "La Navegabilidad en el Curso Alto Del Guadalquivir en Época Romana." *Florentia Iliberritana* 8: 125–147.

Murillo-Barroso, Mercedes, and Marcos Martinón-Torres. 2012. "Amber Sources and Trade in the Prehistory of the Iberian Peninsula." *European Journal of Archaeology* 15 (2): 187–216.

Murillo-Barroso, Mercedes, Enrique Peñalver, Primitiva Bueno, Rosa Barroso, Rodrigo de Balbín, and Marcos Martinon-Torres. 2018. "Amber in Prehistoric Iberia: New Data and a Review." *PLOS One* 13 (8): 1–36.

Needham, Stuart, and Claudio Giardino. 2008. "From Sicily to Salcombe: A Mediterranean Bronze Age Object from British Coastal Waters." *Antiquity* 82 (315): 60–72.

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

Olson, David M, Eric Dinerstein, Eric Wikramanayake, [...], Wesley Wettengel, Prashant Hedao, and Kenneth Kassem. 2001. "Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity." *BioScience* 51 (11): 933–938.

Oxé, August, Howard Comfort, and Philip Kenrick. 2000. *Corpus Vasorum Arretinorum: A Catalogue of the Signatures, Shapes and Chronology of Italian Sigillata.* Vol. 2. Bonn: Habelt.

Özak, Ömer. 2010. "The Voyage of Homo-economicus: Some Economic Measures of Distance." Department of Economics, Brown University mimeo.

Özak, Ömer. 2018. "Distance to the Pre-industrial Technological Frontier and Economic Development." *Journal of Economic Growth* 23 (2): 175–221.

Pandolf, K. B., B. Givoni, and R. F. Goldman. 1977. "Predicting energy expenditure with loads while standing or walking very slowly." *Journal of Applied Physiology* 43 (4): 577–581. PMID: 908672.

Parodi Álvarez, Manuel J. 2012. "La Navegación Interior Ibérica Según Pomponio Mela: Una Visión de la Hispania Romana desde el Fretum-Gaditanum: Ríos Atlánticos Peninsulares. Espacio y Tiempo." *Revista de Ciencias Humanas* 26: 137–156.

Pasquini, Béline, and Christophe Petit. 2016. "Le Portage Entre la Saône et la Moselle dans l'Antiquité (Ier-IVe siècles)." *Environnement et sociétés* 142: 27–32.

Paulsson, B. Schulz. 2019. "Radiocarbon Dates and Bayesian Modeling Support Maritime Diffusion Model for Megaliths in Europe." *PNAS* 116 (9): 3460–3465.

Percoco, Marco. 2015. "Highways, Local Economic Structure and Urban Development." *Journal of Economic Geography* 16 (5): 1035–1054.

Pernicka, Ernst, Bianka Nessel, Mathias Mehofer, and Elvira Safta. 2016. "Lead Isotope Analyses of Metal Objects from the Apa Hoard and Other Early and Middle Bronze Age Items from Romania." *Archaeologia Austriaca* 100: 57–86.

Pernicka, Ernst, Joachim Lutz, and Thomas Stöllner. 2016. "Bronze Age Copper Produced at Mitterberg, Austria, and Its Distribution." *Archaeologia Austriaca* 100: 19–55.

Pétrequin, Pierre, Serge Cassen, Michel Errera, L. Klassen, Alison Sheridan, and Anne Marie Pétrequin. 2012. *Jade: Grandes haches alpines du Néolithique européen, Ve et IVe millénaires av. J.-C.* Centre de Recherche Archéologique de la Vallée de l'Ain.

Pinarelli, Laura. 2004. "Lead Isotope Characterization of Copper Ingots from Sardinia (Italy): Inferences on Their Origins." *Bulletin of the Geological Society of Greece* 36 (3): 1173–1180.

Radivojević, Miljana, Benjamin W. Roberts, Ernst Pernicka, Zofia Stos-Gale, Marcos Martinón-Torres, Thilo Rehren, Peter Bray, Dirk Brandherm, Johan Ling, Jianjun Mei, et al. 2019. "The Provenance, Use, and Circulation of Metals in the European Bronze Age: The State of Debate." *Journal of Archaeological Research* 27 (2): 131–185.

Redding, Stephen J., and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson and William C. Strange, 1339–1398. Elsevier.

Reguera-Galan, Aida, Tania Barreiro-Grille, Mariella Moldovan, Lara Lobo, Miguel A. de Blas Cortina, and José I. García Alonso. 2019. "A Provenance Study of Early Bronze Age Artefacts Found in Asturias (Spain) by Means of Metal Impurities and Lead, Copper and Antimony Isotopic Compositions." *Archaeometry* 61 (3): 683–700.

Reiter, Samantha Scott, Karin Margarita Frei, Heide Wrobel Nørgaard, and Flemming Kaul. 2019. "The Ølby Woman." *Danish Journal of Archaeology* 8: 1–22.

Riley, Shawn, Stephen Degloria, and S.D. Elliot. 1999. "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity." *Internation Journal of Science* 5: 23–27.

Römisch-Germanisches Zentralmuseum in Mainz. "http://www.rgzm.de/samian/."

Sabatini, Serena, and Fulvia Lo Schiavo. 2020. "Late Bronze Age Metal Exploitation and Trade: Sardinia and Cyprus." *Materials and Manufacturing Processes* 35 (13): 1501–1518.

Schauer, Peter, Andrew Bevan, Stephen Shennan, Kevan Edinborough, Tim Kerig, and Mike Parker Pearson. 2020. "British Neolithic Axehead Distributions and Their Implications." *Journal of Archaeological Method and Theory* 27: 836–859.

Scheidel, Walter. 2013. "Explaining the Maritime Freight Charges in Diocletian's Prices Edict." *Journal of Roman Archaeology* 26: 464–468.

Scheidel, Walter. 2014. "The Shape of the Roman World: Modelling Imperial Connectivity." *Journal of Roman Archaeology* 27: 7–32.

Schiavina, Marcello, Sergio Freire, and Kytt MacManus. 2019. "GHS-POP R2019A - GHS Population Grid Multitemporal (1975-1990-2000-2015)." European Commission, Joint Research Centre (JRC).

Shepherd, William R. 1923. *Historical Atlas.* New York: Henry Holt and Company.

Stammann, Amrei. 2017. "Fast and Feasible Estimation of Generalized Linear Models with High-Dimensional k-way Fixed Effects." *unpublished manuscript available at https://arxiv.org/abs/1707.01815*.

Stos-Gale, Zofia A. 2011. ""Biscuits with Ears": A Search for the Origin of the Earliest Oxhide

Ingots." Vol. 9 of *Prehistory Monographs*, 221–229. INSTAP Academic Press.

Stos-Gale, Zofia A., and Noël H. Gale. 2009. "Metal Provenancing Using Isotopes and the Oxford Archaeological Lead Isotope Database (OXALID)." *Archaeological and Anthropological Sciences* 1 (3): 195–213.

Stos-Gale, Zofia A., and Noel H. Gale. 2010. "Bronze Age Metal Artefacts Found on Cyprus-Metal from Anatolia and the Western Mediterranean." *Trabajos de Prehistoria* 67 (2): 389–403.

Talbert, Richard, and Roger Bagnall. 2000. *Barrington Atlas of the Greek and Roman World.* Princeton, N.J.: Princeton University Press.

van Leusen, Martijn. 2002. *Pattern to Process: Methodological Investigations Into the Formation and Interpretation of Spatial Patterns in Archaeological Landscapes.* Rijksuniversiteit Groningen.

Vörösmarty, C. J., B. M. Fekete, M. Meybeck, and R. B. Lammers. 2000. "Global System of Rivers: Its Role in Organizing Continental Land Mass and Defining Land-to-ocean Linkages." *Global Biogeochemical Cycles* 14 (2): 599–621.

Walkowitz, Jürgen. 2003. *Das Megalithsyndrom: europäische Kultplätze der Steinzeit. Beiträge zur Ur- und Frühgeschichte Mitteleuropas*. Beier & Beran.

Wightman, Edith Mary. 1985. *Gallia Belgica.* University of California Press.

Williams, Nicola, and Virginie Boone. 2002. *The Loire.* Melbourne, Paris: Lonely Planet.

Williams, Ross A., and Cécile Le Carlier de Veslud. 2019. "Boom and Bust in Bronze Age Britain: Major Copper Production from the Great Orme Mine and European Trade, C. 1600-1400 Bc." *Antiquity* 93 (371): 1178–1196.

WISE. 2020. "WISE Water Framework Directive Database."