Roman Roads to Prosperity: Persistence and Non-Persistence of Public Infrastructure

Carl-Johan Dalgaard[*]
Nicolai Kaarsen[†]
Ola Olsson[‡]
Pablo Selaya[§]

November 30, 2021

Abstract

In this paper, we explore the link between public transport infrastructure investments made during antiquity and the presence of infrastructure today, as well as the link between early infrastructure and economic activity, both in the past and in the present. We examine the entire territory under dominion of the Roman Empire at the zenith of its geographical extension (117 CE), and find a remarkable pattern of persistence showing that greater Roman road density goes along with (a) greater modern road density, (b) greater settlement formation in 500 CE, and (c) greater economic activity in 2010-2020. Exploiting a natural experiment, we find that both persistence in road density and the strong link between early road density and contemporary economic development weaken to the point of insignificance in areas where the use of wheeled vehicles was abandoned and caravan trade routes replaced road-based trade from the first millennium CE until the late modern period. Studying channels of persistence, we also identify the emergence of market towns during the early medieval period and the modern era as a robust mechanism sustaining the persistent effects of the Roman roads network. Overall, our results contribute to a deeper understanding of the channels through which persistence in comparative development comes about.

Keywords: Roman roads, Roman Empire, public infrastructure, persistence.
JEL classification codes: H41, O40.

[*]University of Copenhagen, CAGE, and CEPR
[†]Danish Economic Council
[‡]University of Gothenburg
[§]University of Copenhagen

1 Introduction

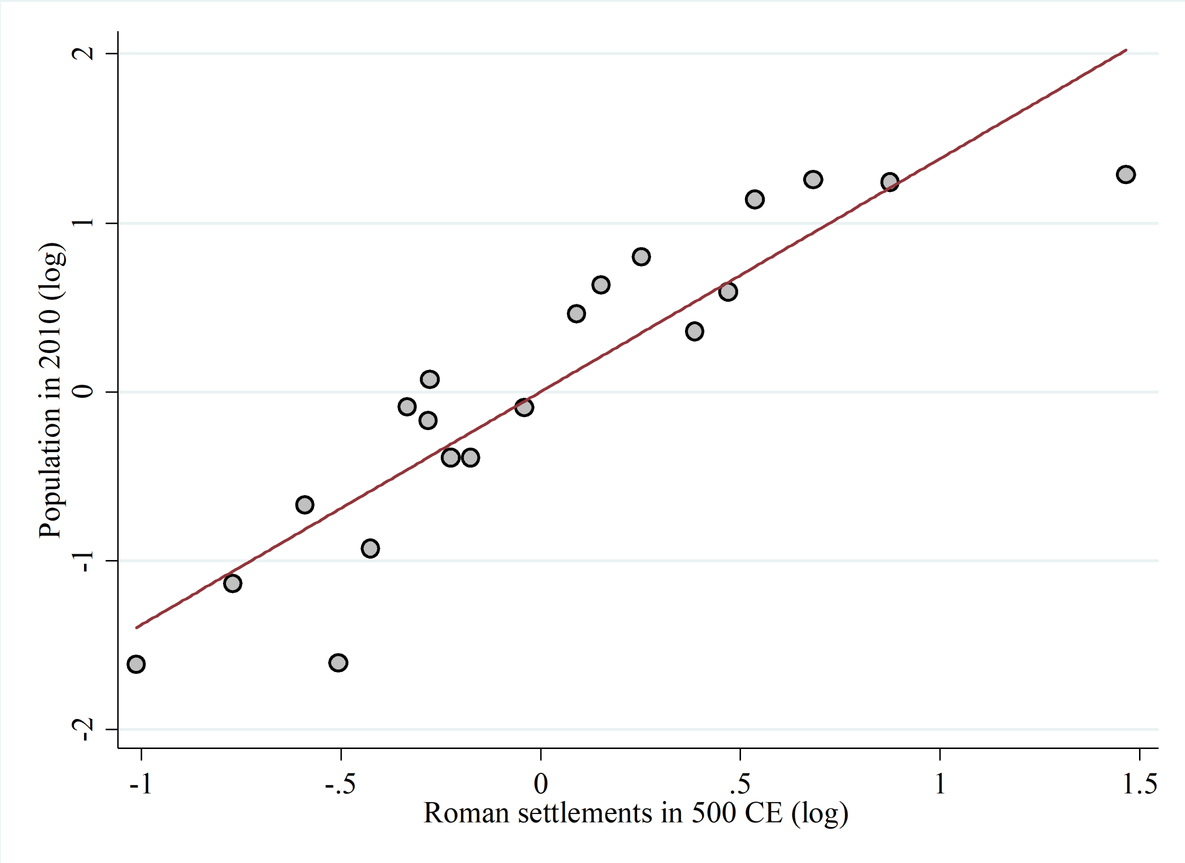
The literature on comparative development has drawn attention to a remarkable degree of persistence in economic activity: places that were relatively more advanced in pre-industrial times tend to be comparatively more prosperous today – see e.g. Olsson and Hibbs (2005), Comin *et al.* (2010), Chanda *et al.* (2015), and Maloney and Valencia (2016). As an illustration, Figure 1 depicts the relationship between contemporary population density and the intensity of Roman settlements in 500 CE, within the territory of the Roman Empire in 117 CE.

This insight has led to extensive research about the fundamental origins of comparative development in (geographic) initial conditions, and historical processes that shaped cultural traits and the institutional infrastructure of societies. Yet much less attention has been devoted to the study of persistence in proximate determinants of growth, or factors that could be the channel that connects fundamentals to economic development in the past and in the present. A deeper understanding of the channels through which persistence in comparative development comes about may provide important clues as to which fundamentals are important, and how to potentially stimulate development in situations where important fundamentals are lacking.

This paper studies the persistence of transport infrastructure, and its effects on comparative long-term economic development, starting in antiquity with the establishment of the Roman roads network. As discussed in detail below, Roman road construction did not follow the rules of infrastructure planning in the contemporary era. The Roman roads were built mainly with a military purpose in mind, and geographic obstacles in the landscape were often surmounted rather than evaded. Despite this, our analysis uncovers a remarkable degree of persistence in road density across time and space: areas that attained greater road density during antiquity are characterized by a significantly higher road density today. Moreover, Roman roads were linked to economic activity by the end of antiquity, and remain a strong correlate of prosperity today. Overall, these results provide evidence that the provision of public transport infrastructure is an important channel through which persistence in economic development (as Figure 1 illustrates) may arise.

In our analysis, we confine attention to localities defined as grid cells of one degree latitude by one degree longitude that were part of the Roman Empire by the beginning

Figure 1: Conditional relationship between population in 2010 and the extent of Roman settlements in 500 CE within the former Roman Empire



Note: The figure shows the conditional binned residual scatter plot of the relationship between population size (in logs) in 2010 and number of Roman settlements (in logs) in 500 CE for 693 grid cells that had some road infrastructure, within the former Roman Empire in ca. 117 CE. The binned scatter plot groups the x-axis variable into equal-sized bins, computes the mean of the x-axis and y-axis variables within each bin, and then creates a scatter-plot of these data points. The underlying regression controls for contemporary country fixed effects and hence estimates the within-country impact of historical Roman settlements on contemporaneous population levels.

of the second century CE (117 CE), and that were *treated* by at least one Roman road. By omitting areas that fell outside the Empire, as well as areas unconnected to the road network, we hope to disentangle the influence of physical infrastructure on economic outcomes from the legacy of Roman rule more broadly.[1]

To further limit the risk of confounding the impact of the Roman roads with Roman influence in general terms, and to filter out the impact of modern day institutions and sub-national differences in cultural values, we also control for country fixed effects as well as language fixed effects throughout our econometric analysis. We believe this strategy makes it unlikely that our results are driven by the legacy of Roman rule in the broadest terms. Nevertheless, a natural remaining concern is that areas receiving more Roman roads may differ in various geographic dimensions that by themselves may have influenced comparative economic development. We attempt to surmount this challenge in two different ways.

For starters, it is important to observe that the risk of confounding the influence of roads with geography may not be as great as one might think. According to the historical literature, it is conventional wisdom that major Roman roads were built in order to facilitate the movement of troops across the empire, rather than with the objective of enhancing economic development. Moreover, the location of roads was arguably only to a limited extent dictated by geographic circumstances. In our analysis we examine these arguments statistically, and find that geographic characteristics account for about a third of the total variation in Roman road density. Hence, we find a claim of orthogonality between geography and Roman road location to be untenable, for which reason we set a rich set of geographical characteristics as a baseline when exploring the persistence of early road infrastructure and its predictive power vis-a-vis economic development over time.

Yet, doubts may still linger about whether that control strategy is sufficient. Therefore, our second strategy to assess the importance of potentially omitted geographic characteristics consists in exploiting the remarkable time span of abandonment of wheeled transportation in North Africa and the Middle East. According to the landmark study by William Bulliet (1990 [1975]), wheeled transport disappeared in North Africa and

[1] See for example Landes (1998) on the legacy of Roman rule. Recent research has also shown that areas under Roman influence arguably developed different institutions from areas outside Roman direct influence (e.g. Glaeser and Shleifer, 2002).

the Middle East somewhere between the fourth and sixth centuries CE. Eventually wheeled transport vehicles had to be reintroduced with the ascent of the automobile.[2] Consequently, following the fall of the western part of the Roman Empire, the roads fell into disrepair, and ultimately went out of use in North Africa and the Middle East. In contrast, Roman roads continued to be maintained and in use in Europe after the fall of the Western Roman Empire (Glick, 2005 [1979]; Hitchener, 2012; Wahl, 2017).[3]

From the point of view of our study, this natural experiment has two important implications. First, as the ancient roads fall into disuse and thus are left unmaintained, it becomes much less likely that modern roads are built in their place. In that sense, one would expect to see a far weaker link between Roman road density and modern road density outside the European part of the empire. Second, if the roads create persistence in economic development, one would expect to find a far weaker link between ancient infrastructure and modern-day economic activity within the regions where roads temporarily lost relevance. Consistent with these conjectures, we find that there is no significant link between ancient infrastructure and modern infrastructure *within* North Africa and the Middle East. Moreover, within these two regions the ancient infrastructure is not a significant predictor of economic activity today. In contrast, in Europe – the region where roads continued to be used and therefore maintained – ancient roads predict modern roads as well as prosperity.

This differentiated effect of Roman road density is revealing from the point of view of identifying channels of influence. Naturally, the fundamental principles governing the construction of the Roman roads were the same throughout the Empire. If our baseline results are tainted by omitted variables bias – for example due to missing geographic characteristics that matter both for road location and subsequent economic activity – one would expect to see evidence of an apparent persistence of infrastructure density as well as a persistent impact from ancient infrastructure on modern economic activity throughout the entire empire. Accordingly, in light of our findings, the "abandonment of the wheel"-experiment provides fairly compelling evidence of the mechanism under scrutiny: persistence in public infrastructure leads to persistence in economic activity.

[2] See also Chaves *et al.* (2014) on the absence of wheeled transport in Sub-Saharan Africa and its introduction during the early colonial period.

[3] Even if the roads in some parts of Europe were left unmaintained, they may have left a mark in the landscape, which is less likely to arise in the MENA. See Section 4.3.1 for further discussion.

To further examine the Roman roads pattern of non-persistence in the Middle East and the North of Africa (MENA), we look at the relationship between roads and the caravan routes that emerged in this region with the rise of Islam. This exercise shows a mix of statistically null and small but negative connection between Roman roads and the caravan routes that emerged in MENA after 632 CE. However, this is an interesting result by itself, as it suggests that caravan trade routes that replaced road transport in this region may not have exhibited persistence or had persistent economic effects in the same way as road infrastructure, and therefore helps to understand better the drivers of non-persistence in comparative development.

In the last part of our empirical analysis, we analyze a potential mechanism behind the persistence from ancient to modern roads in Europe: the emergence of market towns during the extended medieval period. Using data from a geographical corridor along the Roman *limes* in modern Germany, we show that in the former Roman areas that had a dense road network in antiquity, market towns were significantly more likely to arise during the medieval period.

The present study is related to several strands of literature. First, it is related to research on long-run persistence in economic development – see recent summaries e.g. in Spolaore and Wacziarg (2013), Nunn (2014) and Ashraf and Galor (2018). This literature has largely been concerned with the influence and origins of fundamental determinants of productivity that ultimately can explain persistence in comparative development.[4] In contrast, the present paper focuses on channels of persistence, or the persistence of *proximate* sources of growth. In particular, the focus is on whether public goods provision appears to be persistent over long periods of time. From this perspective, our work is related to Chen, Kung and Ma (2020), who look within China and document a remarkable persistence of another proximate source of growth: education. The authors argue persuasively that the observed persistence is (at least in part) explained by the emergence of a pro-education culture prompted by early educational investments. Importantly, the argument is supported by the fact that persistence in schooling weakens markedly in areas that were particularly exposed to the anti-intellectual Cultural Revolution. Huilery (2009) illustrates similar long-term effects for the case of colonial investments in education, health, and infrastructure in West

[4] Recent contributions include Galor and ”Ozak (2016), Gorodnichenko and Roland (2017) on culture and development; Angelucci *et al.* (2017) on institutions, and Andersen et *et al.* (2016) on geography.

Africa; and Franco, Galiani and Lavado (2021) the long-term effects of ancient transport infrastructure in South America – they show in particular that roads built during the Inca Empire in the territory of present-day Peru are associated with higher wages, better nutrition and better schooling outcomes today and, again importantly, that the provision of primary schools and roads is an important driver of the persistent effects.

Our study documents that the provision of Roman public road infrastructure is similarly persistent over very long periods of time, except in areas where roads lost economic value early on. By carefully controlling for the influence of (countrywide) formal institutions and (within-country) informal institutions, our findings indicate that this persistence is unlikely to be mediated by the emergence of cultural values or institutions. Instead, our analysis suggests that later roads likely were built on top of older roads, thus creating persistence in road density.[5] Of course, in places where the signs of the early roads disappear, modern roads would be less likely to be located along the ancient trajectories.

Second, our paper also contributes to a small but growing literature on the economics of the Roman period (Finley, 1973; Temin, 2006; Bowman and Wilson, 2009; Michaels and Rauch, 2017; Benedictis *et al.*, 2018). A striking feature of the classic work by Finley (1973), for instance, is its relative neglect of the general importance of roads. A common theme is that road transport was inferior to shipping in terms of efficiency and hence of less importance. Recent research however has started to re-examine the influence of the Roman road network on long-run development. Bosker *et al.* (2013), studying determinants of city growth between 800 CE and 1800, document that cities located at intersection points (hubs) between Roman roads grew bigger in Europe, but not in the Middle East and North Africa. The authors note that this was most likely due to the emergence of more efficient camel transportation in the extensive desert areas of that region, which followed very different paths than the Roman roads as we describe above. Bosker *et al.* (2013) also study the impact of caravan routes and hubs in the Muslim world. [6] The present study focuses on the persistence of infrastructure over time, which is not discussed in Bosker *et al.* (2013). Three of our outcome variables are

[5] Our finding of persistence in road investments over time is also consistent with the presence of local increasing returns to scale in infrastructure investments, as emphasized in a literature inspired by Krugman (1991).

[6] On the link between city location and Roman roads within Europe, see also Bosker and Buringh (2017).

further contemporary whereas Bosker *et al.*'s empirical analysis ends in 1800.

Using data on excavated Roman ceramics, Flückinger *et al.* (2021) study network connectivity within the general Roman transport system. The authors find that trade between locations during antiquity still explains a great deal of the variation in the spatial pattern of contemporary firm ownership. A key mechanism for this persistence is the convergence in preferences and cultural values. The analysis in Flückinger *et al.* (2021) is based on a gravity-type of empirical analysis, using measures of bilateral integration between regions as the dependent variable, whereas our analysis uses the cross-sectional variation within countries and language areas with night lights intensity and levels of population as proxies for levels of contemporary economic prosperity.[7]

Also related is the work of Wahl (2017), who investigates the Roman *limes* inside Germany and finds that contemporary development is more advanced on the old Roman side of the *limes* border using a regression discontinuity design. Wahl (2017) identifies the road system as an important explanatory factor, which is consistent with our findings about the European part of the Empire. At the same time, our analysis provides evidence on how shocks in the past, such a the abandonment of the wheel, can importantly perturb development trajectories with long-run implications for comparative development.[8] Using Wahl's (2017) sample of grid cells along the Roman border within modern Germany, and combining it with data on the rise of market towns after 850 CE from Cantoni and Yuchtman (2014), we show that the more prevalent emergence of markets in the old Roman areas suggests that the development of trade nodes was a key intermediate channel for the persistence of historical infrastructure into modern times.

Third, in recent years, a surge of interest in the economic effects of infrastructure has resulted in a number of important studies. The results indicate that infrastructure investments often have a strong positive influence on population growth and economic activity.[9] A major difference with the present study obviously consists of the time

[7]Using transcribed text from ancient tablets, a recent paper by Barjamovic et al (2019) also use a structural gravity model to estimate the geographic location of lost cities in Old Assyria, 2000-1650 BCE.

[8]Relatedly, McGrath (2016) explores the correlation between Roman roads and nightlights in current day Scotland. As our sample ends at the Wall of Hadrian, which is essentially the border of Scotland, McGrath (2016) can be seen as providing external validity to our reduced form results.

[9]On roads see Fernald (1999), Michaels (2008), and Bird and Straub (2015). On railways see Banerjee *et al.* (2012), Donaldson (2012), Jedwab and Moradi (2015), and Hornung (2015).

horizon over which the consequences of infrastructure investments are assessed; with a two millennia perspective we believe the present study has the longest observation window hitherto explored. At the same time it is important to stress that the objective of the present study is not to estimate the productivity gains from infrastructure per se. In fact, our approach does not allow us to distinguish whether areas that received more infrastructure investments outgrew areas with less infrastructure because of productivity benefits from public goods, or if more public goods in a particular location simply attracted activity from other locations. As should be clear, the main focus of the present study is rather to assess the persistence of public goods provision and thereby whether public goods provision has a significant role to play in the observed pattern of persistence of comparative development.

The paper proceeds as follows. In section 2, we present our data on Roman roads, and our outcome variables. In section 3, we outline an historical background on the assignment of Roman roads as well as formal tests of the geographical determinants of Roman road density. In section 4, we present and discuss the main empirical results. Section 5 concludes.

2 Data

In this section we describe the unit of observation, the central independent variable for the regressions that follow (the Roman roads variable), and the main dependent variables. The appendix contains a description of our control variables, as well as summary statistics.

2.1 Unit of observation: Grid cells

We study the area within the borders of the Roman Empire in the year 117 CE, which is around the time when the empire attained its maximum territorial expansion.[10] To carry out an analysis at the local level, we divide the Roman Empire area in grid cells of 1x1 degrees of latitude by longitude. In places where these grid cells are cut by the borders of the Roman Empire, or crossed by a country or a language area border, we

[10]Data on the extent of the Roman Empire in the year 117 CE are from the Ancient World Mapping Center at the University of North Carolina at Chapel Hill, http://awmc.unc.edu.

split them following the corresponding borderline. This approach allows us to control for country and language fixed effects throughout the entire econometric analysis. As a robustness check, we also present results with smaller (0.5x0.5) grid cells.

2.2 Independent variable: Roman roads

The raw data for the Roman roads come from a digitized map of the road network illustrated and documented in Talbert's (2000) "Barrington Atlas of the Greek and Roman World".[11] We focus on all roads identified in the digitized map as being of major importance, and drop all roads identified as minor because of the difficulties involved in getting precise traces of small ancient roads.[12].

To construct a measure of the influence that Roman roads have had, we draw areas (buffers) of 5 km around each road in the network, and compute the percentage of area of these buffers within all our grid cells. As an illustration of this procedure to measure the degree of influence of Roman roads, Figure 2 shows the 5 km buffer around all major roads in the Roman Empire, and Figure 3 zooms in on the road system and the buffer around the main Roman roads crossing the area of Lutetia (contemporary Paris) in France and Londinium (today London) in England.

At first sight, a 5 km buffer may seem large, and it is certainly possible to construct a smaller buffer. The choice is made so as to accommodate the fact that early Roman infrastructure was associated with adjacent investments (for example drainage systems). In order to be sure to envelope the total treatment resulting from the road construction, we use the said buffer size. This still allows a lot of variation since our unit of observation is an area of 5,226 sq km on average (see table with summary statistics in the appendix).

As mentioned in the Introduction, we confine our attention to grid cells that were *treated* by at least one Roman road. That is, we focus the analysis exclusively on the

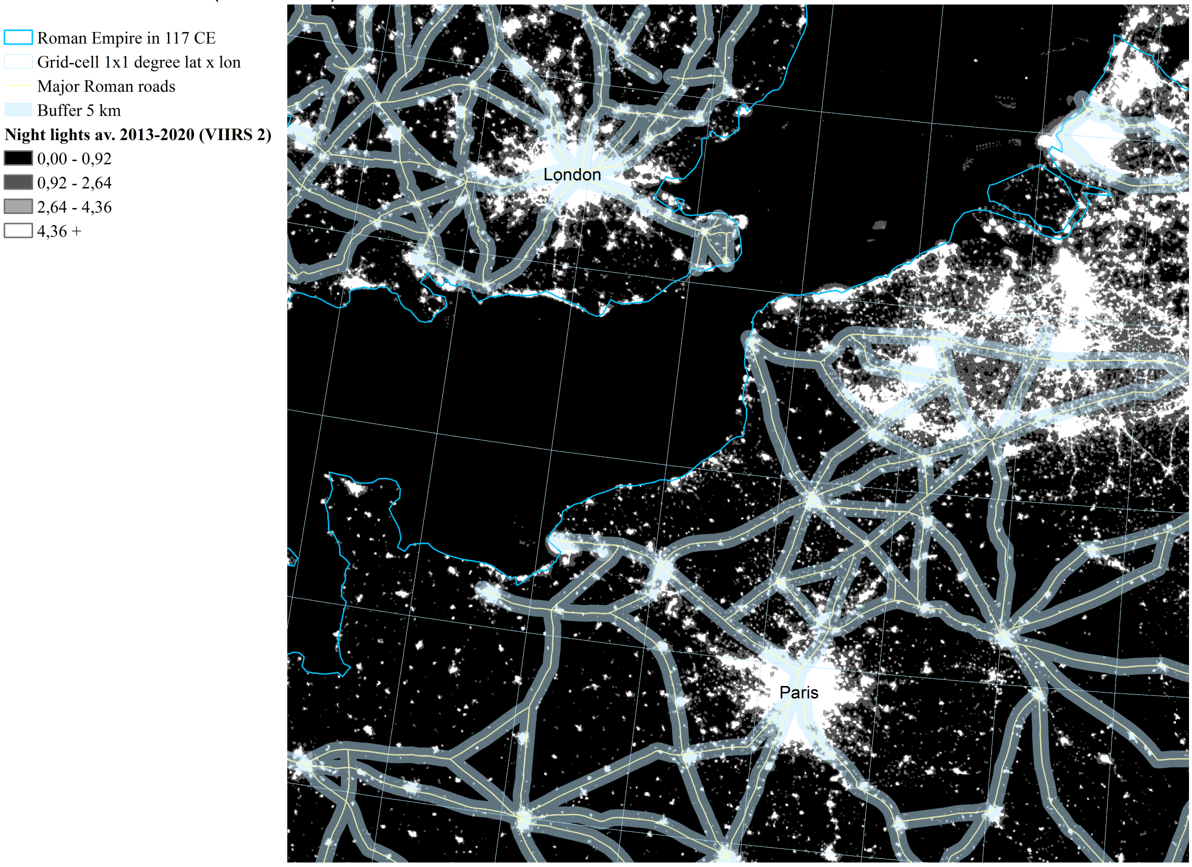
[11]The digitization was done by McCormick *et al.* (2013), as part of the Digital Atlas of Roman and Medieval Civilization (DARMC) project of the Center for Geographic Analysis at Harvard University, http://darmc.harvard.edu.

[12]The Barrington Atlas notices, in fact, that the roads system is perhaps "the most difficult element to map". Page 262 in the map-by-map directory that accompanies Talbert's (2000) Barrington Atlas describes, for example, the difficulty of mapping minor roads that had neither milestones nor paving in the outline of Augustan roads or in routes through the Alpine valleys (Map 18). Page 169 describes, as another example, the use of tree-ring dating methods to overcome the difficulties in tracing parts of the Via Claudia Augusta (Map 12). See http://press.princeton.edu/B_ATLAS/B_ATLAS.PDF for a detailed description of all roads and other features contained in the Barrington Atlas.

Figure 2: Roman roads and contemporary night light intensity among 1000 grid cells within the Roman empire in 117 CE

Note: The map shows major Roman roads within the boundaries of the Roman Empire in 117 CE with average nightlights intensity between 2013 and 2020 in the background.

Figure 3: Roman roads and contemporary nightlights intensity around Lutetia (Paris) and Londinium (London)



Note: The figure shows the rectangular 1x1 degree latitude x longitude grid cells including the hub of Roman roads around Lutetia (contemporary Paris) and Londinium (contemporary London). Buffer zones of 5 km on either side are shown around the roads. Background colors indicate the strength of modern (average 2013-2020) nightlights intensity.

intensive margin of infrastructure investments. Also, in order to disentangle the influence of the physical infrastructure on economic outcomes from the legacy of Roman rule more broadly, we drop areas that fell outside the Empire or were unconnected to the road network within the Empire. (However, in the interest of completeness, we also report, in the appendix, results where we measure Roman roads along the extensive margin, within the borders of the Roman Empire.)

2.3 Dependent variable I: Modern roads

Data for modern roads are taken from ESRI's World Roads basemap layer package, a comprehensive and global dataset containing major roads and ferries routes around the world. The dataset is administered by ESRI Data and Maps, and was created in 2011 and updated in 2021.[13]

Just as we do for the Roman roads, we build our indicator of the intensity of modern roads by constructing a buffer of 5 km around the network of major roads, and computing the percentage of the buffer area within each grid cell of 1x1 degrees of latitude and longitude, within each country and language area, within the contours of the Roman Empire in the year 117 CE.

2.4 Dependent variable II: Roman settlements

As a measure of economic development at the end of antiquity we use the number of major Roman settlements in the year 500 CE. We take these data from the Digital Atlas of the Roman Empire (DARE) constructed by Åhlfeldt (2017).[14] The data set was compiled using Talbert (2000) and other sources, and contains the location of settlements, mines, forts, villas and various other localities. The observed time span of existence is indicated for each locality. We compute the number of major settlements that existed in year 500 CE within each grid cell.

[13]The data are available at https://www.arcgis.com/home/item.html?id=83535020ce154bd5a498957c159e3a99.

[14]The DARE data is continuously updated. Our version was downloaded from the Pleiades data site https://pleiades.stoa.org/home on August 16, 2017.

2.5 Dependent variable III and IV: Nightlights and population density

As proxies of local economic activity today, we rely on the average intensity of lights at night between 2013 and 2020, and the level of population density in 2010.

The raw data for lights comes from nighttime satellite imagery collected by the Earth Observation Group (EOG) at the Colorado School of Mines, which are available as at high granularity and monthly and yearly frequencies. We use the most recent EOG's data available, which are the Visible and Infrared Imaging Suite (VIIRS) Annual VNL 2 time series of annual masked averages of nighttime intensity (which are averages computed with background and noise set to zero values). These data are available at a granularity of 15 arc seconds (about 500 m around the Equator), for the period 2013 to 2020.[15] We compute the 2013-2020 average of the raw data within each 1x1 grid cell within the contours of the Roman Empire in 117 CE, and use it as our main indicator of nighttime light intensity. Figure 2 and 3 above show the average 2013-2020 geographical distribution of nightlights around the area covered by the Roman Empire in 117 CE.

For population density we use the UN-adjusted 2010 population count from the Gridded Population of the World version 4 database, that adjusts gridded population numbers to United Nations (UN) estimates of national population counts.[16] To construct population densities, we simply sum population numbers within each grid cell and divide the sum by the total grid cell area.[17]

3 The Roman roads

3.1 Historical introduction

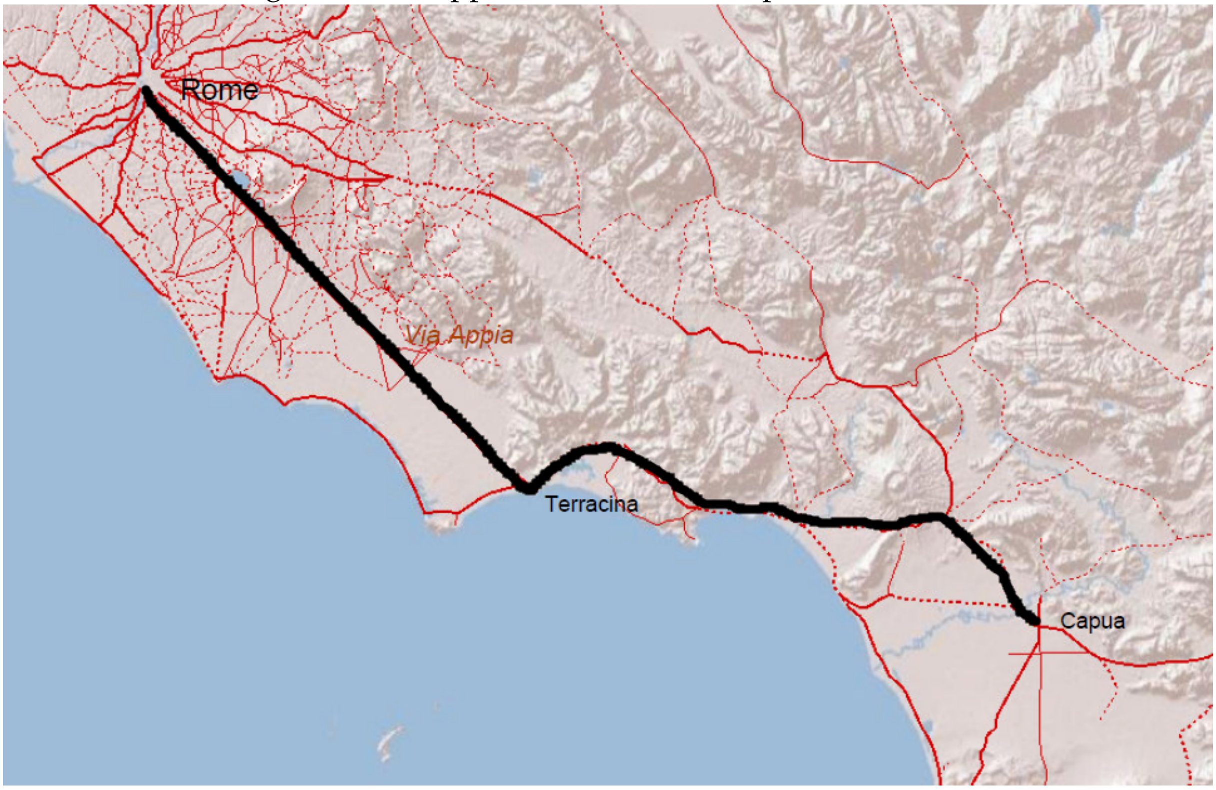
The Roman road construction program during antiquity is generally considered to have been initiated in 312 BCE when censor Claudius Appius started the construction of a paved, all-weather road, subsequently named *Via Appia*, from Rome to Capua

[15] The data are available at https://eogdata.mines.edu/products/vnl.

[16] The alternative are unadjusted population levels, which are based on individual countries censuses and population registers, which we avoid with the aim of having more comparable data.

[17] We use the finest resolution data of 30 arc seconds to compute our variables. The raw data are available at http://sedac.ciesin.columbia.edu/data/collection/gpw-v4.

Figure 4: Via Appia from Rome to Capua in 312 BCE



Note: The thick black line shows Via Appia whereas the red lines show other, later Roman roads. Source: Created on the basis of data in Talbert (2000).

(Figure 4). The immediate reason for the construction of Via Appia was the ongoing Second Samnite War in which the Roman armies were trapped around Capua due to shortage of supplies from Rome. It is believed that the road was completed in 308 BCE. With the new and more efficient supply lines, the Romans defeated the Samnites in 304 BCE and Via Appia was eventually extended all the way to the southeastern port city of Brundisium a few decades later (Laurence, 1999).

Via Appia was neither the first road in the Mediterranean area (the Persians under Darius the Great had for instance constructed extensive royal roads in the 5th century BCE), nor in Roman territory (earlier, non-paved roads are mentioned by ancient Roman sources). Nonetheless, it would come to serve as a model for future road constructions, first on the Italian peninsula and later in the broader empire.[18] At the peak of the Roman empire at the death of Trajan in 117 CE, it is estimated that the empire hosted about 80,000 km of paved road (Gabriel, 2002). As Figure 2 shows, the road

[18]In Italy, for instance *Via Flaminia* (completed in 220 BCE) connected Rome with the Adriatic coast, whereas *Via Aemilia* (187 BCE) cut through the Po plain and made that imporant agricultural area available for Roman colonization.

system connected regions in current Britain, Western Europe, Eastern Europe, North Africa and the Near East.

Because of their military purpose for achieving effective Roman control of a territory, the construction of these public highways (*viae publicae*) was carried out by Roman legions, and it was typically commissioned by a censor and administered by curatores in Rome.[19] Most of these roads were paved with stone and cement. Road building also included supporting public goods such as bridges, tunnels, guest houses, and drainage systems which required substantial engineering skills. Scholars have suggested that the construction of roads also fostered the use of ground surveys and maps (Davies, 1998). Ordinary citizens sometimes had to pay tolls at city gates and bridges, and the military always had priority. The *viae publicae* network was complemented by local roads, *viae vicinales*, which typically linked the major roads to a town or to other major roads. These roads were mainly the responsibility of local governments (Laurence, 1999).

3.2 The assignment of Roman roads: Historical priors

There are three main reasons why the Roman road construction program almost presents itself as a natural experiment, from the vantage point of the historical literature: i) The *military purpose* of the roads, ii) the preferred *straightness* of the construction, and iii) their construction in *newly conquered* and often undeveloped regions.

First, just as mentioned above with the early experience with the Appia during the war with the Samnites, the purpose of the roads was to increase the speed and the ease with which the legions could reach locations of military interest – including territories of ongoing campaigns, army bases and Roman colonies that provided the army with essential supplies. Viae publicae also played a key role for the consolidation of power and hegemony in newly conquered areas.[20] When the Roman limes stabilized along its northern and eastern frontiers, the road system was used to transport marching troops to the their legionary bases along the border. Very soon, the roads were also

[19]Censor and curator were public offices in Republican Rome. During Imperial times, road construction was mainly carried out by the emperors.

[20]Laurence (1999) argues that a broader objective with the roads was also to demonstrate the general technological superiority and political commitment of the Romans to the peoples in neighboring areas. The construction of roads signalled an ability to even change the geography of landscapes, which presumably greatly impressed many of Rome's rivals.

used by traders and for transportation of agricultural goods, but this was not the main intention.

Second, Roman roads were typically very straight over extensive distances. The ambition of the road engineers was typically to connect an existing point A in an area under Roman control with a specific point B in an area where power was less consolidated. The example of contemporary Rimini and Piacenza in Figure A1 in the Appendix illustrates this tendency. An obvious reason for choosing a straight road was the shorter distance and the lower costs in terms of building material and soldier efforts.[21] The straightness of the roads implied that they often passed right over hills and across difficult terrain. Via Appia, for instance, passed like a straight line right from Rome to the existing colony in Terracina through the Pontine marshes (Figure 4). Malaria was prevalent in this area and the Romans had to construct drainage systems in order to be able to get through. The marching armies were not necessarily much constrained by these difficult conditions, but it has been claimed that the steepness of the roads often made them unsuitable for commercial ox-drawn carts with agricultural goods (Mokyr, 1990). The fact that the major roads tended to be straight also suggests that in between the two connecting points A and B, the highway system often was not adjusted to take into account pre-existing local economic or other social characteristics. For instance, the Romans consciously avoided linking Via Appia to a number of existing Latin settlements in the vicinity (Laurence, 1999).

This relates to the third argument, namely that the Roman roads were often constructed in newly conquered areas without any extensive, or at least not comparable, existing network of cities and infrastructure. The Roman roads were laid out in territories in which they had limited prior knowledge and where they had the aim of quickly securing Roman hegemony.

As an illustration of this point, consider the case of Lugdunum (contemporary Lyon). Julius Caesar's conquest of Gaul north of the Mediterranean coast was completed relatively rapidly during a frantic campaign in the 50s BCE. There were many existing towns and cities in Gaul when the Romans arrived, but there was not a state in any sense comparable to the Roman polity, and most scholars refer to Gaul as proto-urban at the time (Woolf, 1998). In year 47 BCE, Caesar created a Roman colony in

[21]See Davies (1998) for an account of how the Romans managed to keep the roads straight between two points without modern surveying tools.

the important town of Vienne, 30 km south of contemporary Lyon in the Rhone valley and, at the time, the main settlement (referred to by Caesar as *oppida*) of the Gallic Allobroges tribe (see Figure A2 in the Appendix). In 43 BCE, the Romans were expelled from Vienne by the Allobroges. According to the Roman historian Dio Cassius, the Roman Senate then ordered the governor of Gallia Transalpina to found a new city for the refugees from Vienne to the north at the intersection of the Rhone and Saone rivers. This city became the Roman town of Lugdunum. According to Åhlfeldt (2017), this location was not an important existing oppida.

Shortly after the establishment of Lugdunum, Marcus Vipsanius Agrippa, the governor of Gallia Transalpina, initiated an extensive road building project in order to consolidate Roman rule in Gaul. Lugdunum was connected southwards along the Rhone to the important cities of Vienne, Avignon and Massilia. Agrippa also built extensive roads towards the Atlantic to the west, towards the North Sea, and towards the Rhine to the east, thereby making Lugdunum a key hub in Roman Gaul (Gros, 1991). The city experienced a rapid growth as a result and soon eclipsed even the old Greek colonies to the south. It became the capital of and gave name to the Roman province of Gallia Lugdunensis, and served as the primary Roman city in Gaul for more than two centuries.

The example indicates that the Roman decision to make Lugdunum a hub of road construction was probably a combination of good geographical fundamentals (the Rhone and Saone intersection), the historical accident related to the hostility to Romans in the previously much more significant town of Vienne, and the need to quickly consolidate power in Gaul. We do not have strong reasons to believe that Roman road construction was based on an already existing network of prosperous towns in the area. Figure A2 shows the pre-Roman oppida in the Lugdunum area, as well as the subsequent Roman roads and settlements. At least around this key city, there are no indications that the Romans consciously tried to connect to older settlements.[22]

[22]Michaels and Rauch (2016) find however that the location of existing oppida – pre-roman fortified towns – does seem to predict the location of Roman towns in Gaul.

3.3 The assignment of Roman roads: A formal test

In Table A3 in the Appendix, we investigate the determinants of Roman road density. As described in Section 2.1, our units of analysis are grid cells of 1x1 latitude-longitude degrees within the borders of modern countries and territories covered by the extent of the Roman empire in 117 CE. The dependent variable is (log) Roman road density (or the percentage of a 5 km buffer around the Roman road system that lies within the total area of a grid cell). Only cells featuring at least a part of a road are in the sample. We will thus throughout focus on the intensive margin of Roman road construction.

The question we examine in this part is essentially the extent to which the received perception from the historical literature, suggesting a limited influence of geography and pre-Roman development on road investments, is accurate. Naturally, if geography does not play a significant role for road assignment, it lessens the need to control for it in the regressions to follow.

The estimates in Table A3 suggest that Roman road construction was more intense on lower than on higher elevations, relatively close to coastlines, major rivers, and Roman borderlines, and where caloric potential before 1500 CE was relatively large. Perhaps surprisingly, road construction was more dense further away from natural harbours. In the smaller sub-samples where we have proxies for pre-Roman development (oppidas, in the case of Europe, and the timing of the Neolithic), we find very little evidence that such factors influence road density (columns 5 and 6). Even areas featuring mining activity during Roman times are not characterized by greater road density (column 4). When all factors are aggregated in column (7), we find that our total set of geographical covariates explain about a third of the total variation in Roman road density.

Table A3 thus shows that geography needs to be controlled for when examining the persistence of roads and the link between ancient road density and economic activity today. Since geographic features naturally may influence economic development and the location of modern roads in their own right, they are essential controls in the remaining.

4 Ancient roads, modern roads, and economic activity

4.1 Empirical specification

We take the following cross-sectional specification to the data:

$$\log\left(y_{grc}\right) = \delta_c + \delta_r + \beta \cdot RRD_{grc} + \mathbf{X}'_{grc}\mathbf{fl} + \epsilon_{grc}. \tag{1}$$

Our dependent variable, pertaining to grid cell g in language region r and country c is denoted y_{grc}. The dependent variables of interest are, respectively, (log) modern day road density and (log) economic activity during antiquity and today. In the latter case we employ both nightlights (following Henderson *et al.*, 2012) and population density (e.g., Rappaport and Sachs, 2003) in 2010. The independent variable of particular interest is log (1+) Roman road density, RRD_{grc}.[23]

In an effort to control for countrywide institutions we include a full set of country fixed effects, δ_c. In addition, since past research has documented important within-country variation in culture that affect economic activity (e.g., Tabellini, 2010; Michalopoulos and Papaioannou 2013), we rely on a full set of language fixed effects as a proxy, δ_r, following Andersen *et al.* (2016).

$\mathbf{X}_{grc}$ contains additional controls, which can broadly be partitioned into three categories. First, geographic variables that involve latitude, longitude, ruggedness, elevation and controls for soil quality. Second, proximity to water bodies which involves distances to coast, major rivers and natural harbors. Third, a set of variables that control for distances to the border of the empire and the current capital. In addition, we also control for the location of historical mines, as well as distances to coal resources, medieval universities and to Wittenberg (as a proxy for Protestantism). Finally, we control for grid cell (*grc*-cell) area in all specifications, as it varies with latitude and longitude due to the earth's curvature, modern country limits, language areas, and the borders of the former Roman Empire.

In terms of statistical inference, we take in consideration two main aspects. First, Abadie *et al.* (2017), argue that cluster adjustments for standard errors should only be performed if there are strong theoretical priors to do so. In particular, they argue that

[23] Accordingly, since $\log(1+x) \approx x$ the coefficient β strictly speaking has the interpretation of a semi-elasticity.

clustering is only relevant to address an experimental design issue and/or a sampling design issue. Briefly, the former issue arises if the treatment focus occurs at a higher level of aggregation than the unit of observation, whereas the second one emerges if multi-level sampling is taking place (e.g., first in a sample of countries and then in a sample of regions within those countries). In the present case our sample consists of all the grid cells within the Roman Empire that where treated by Roman roads, which means neither of the two issues arises. Accordingly, we present standard errors that are robust to heteroskedasticity throughout the empirical analysis in the paper.

However, spatial correlation is still a natural concern in persistence and historical studies - see e.g. Kelly (2021) who argues that spatial auto-correlation in that types of studies often leads to inflated t-statistics that complicate interpretation of estimated results. In our case this is also a reasonable concern because roads are necessarily connected across space, and the level of road density in one grid cell will always be correlated with density in at least one neighboring grid cell. Similarly, we can suspect from visual inspection that some of our outcome variables, such as nightlight intensity as depicted in Figures 2 and 3, are also spatially correlated since nightlight intensity in one cell will often be correlated across neighboring cells. In order to address the possibility of an arbitrary form of spatial clustering in our main variables, we follow the approach suggested by Colella *et al.* (2020), and proceed in two steps.

First, we test for spatial correlation in both the main outcome variables and the main treatment variable, by examining their spatial correlograms – which display Moran's I for different distance bands. We find that the degree of spatial correlation is significant within a band of 0 to 10 neighboring grid cells (or 0 to 72.3 km, evaluated at the mean) for all our main variables. We also find that the degree of spatial auto correlation in all cases decays significantly in the band of 0 to 20 or 25 grid cells (0 to 1446 or 1807 km), and that it disappears statistically in the band of 0 to 30 or more grid cells (0 to 2169 km or more).[24]

In presence of spatial clustering, "[...] the quality of inference tends to improve with the inclusion of additional control variables that exhibit a spatial kernel comparable to the one of the outcome variable and/or that of the regressor of interest" (Colella *et al.*, 2020, p. 16). Given that our regressions include a variety of controls that also tend to

[24] Appendix Figures A3 and A4 show for example the spatial auto correlogram for nightlights intensity and Roman roads.

be spatially correlated (for instance country fixed effects, language area fixed effects, and elevation), as a second step we check for the presence of spatial correlation in the residuals of our main (richest) regression specifications. We find that these residuals exhibit a much lower degree of spatial auto correlation, which decays fast and disappears after distance bands of 0 to 3 or 5 grid cells (0 to 217 or 361 km).[25] Therefore, taking 5 grid cells or 361 km as the maximum distance threshold, we compute Conley (1999) standard errors that account for spatial correlation within that distance, and report them for all our main results.

4.2 Baseline results

In Table 1, we explore the link between road density during antiquity, and road density today, in our full sample. The specifications in the tables ahead always follow the same pattern with only basic geographical controls in the first column, then introducing country and language fixed effects as well as sets of related control variables, with all controls included together in column 7. Robust (as well as Conley) standard errors are reported for the main variable Roman roads.

Table 1

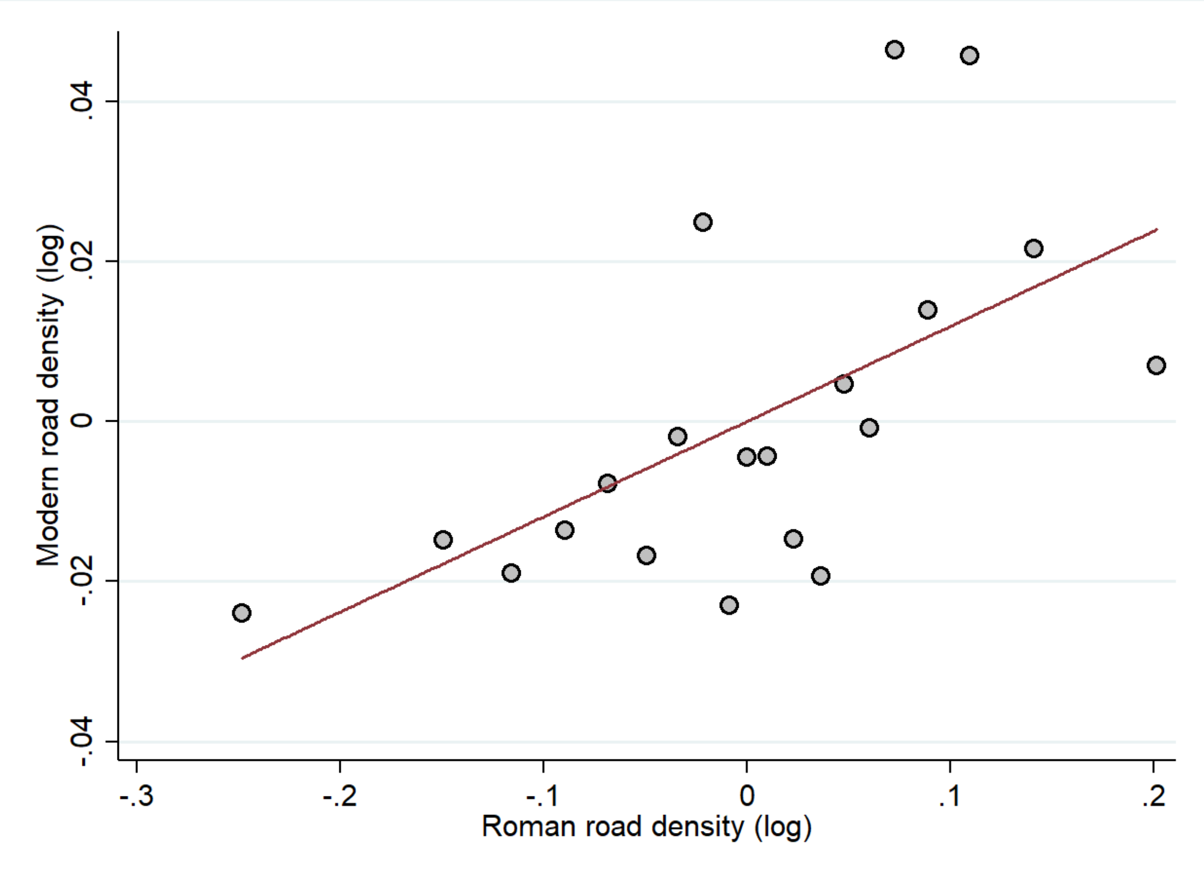
As column 1 of Table 1 shows, ancient roads and area, latitude and longitude controls account, on average, for about 23 percent of the current differences in modern-day road density within our sample. At the level of raw partial correlation, one percent higher Roman road density is associated with an increase in modern day road density of about 0.33 percent.

In column 2 we introduce country fixed effects and in column 3 we introduce simultaneously country fixed effects as well as language fixed effects, which allows us to partial out the influence from institutions and cultural value variation within nations. The economic significance of Roman road density declines, but only to a minor extent.

Adding the first set of geographic controls, involving e.g. a measure of post-1500 caloric potential, makes more of a difference. Collectively, the controls add a lot in

[25] Figure A5 in the Appendix displays, for example, the spatial correlogram for the residual of a regression of nightlights on Roman roads, controlling for country-language fixed effects and all our other baseline controls included.

Figure 5: Conditional relationship between modern road density in 2000 and Roman road density in 117 CE within the former Roman Empire



Note: The figure shows the conditional binned residual scatter plot of the relationship between modern road density (in logs) in 2000 and Roman road density (in logs) in 117 CE for 693 grid cells within the former Roman empire. The binned scatter plot groups the x-axis variable into equal-sized bins, computes the mean of the x-axis and y-axis variables within each bin, then creates a scatter plot of these data points. The underlying regression in Table 1, column 7 controls for the full set of geographical controls, as well as for country and language fixed effects, and hence estimate the within-country impact of historical Roman settlements on contemporaneous population levels. See text and Appendix for exact variable definitions.

explanatory power, and reduces the elasticity of Roman roads to about 0.11. As seen from the rest of the columns, the apparent persistent positive influence of ancient infrastructure on modern infrastructure remains when we add further controls, and all of the controls collectively, in column 7. Standard errors increase, however, and the level of significance falls in the last column. Conley standard errors are usually somewhat higher than robust standard errors but do not change the inherent p-values of the estimates. In total, historical road networks and all the controls together account for about one half of the variation in contemporaneous road network across grid cells (R-squared = 0.52).

Figure 5 shows the partial correlation in the model estimated in Table 1, column 7, using a binned partial residual plot in order to illustrate more clearly the results. Consistent with our estimation approach, we use a linear fit to summarize the relationship between Roman roads and modern roads.[26] The positive link appears well determined.

In Table 2 we turn to the link between Roman roads and economic activity around the collapse of the Western Roman Empire at the end of the fifth century. As a measure of economic activity we use the density of major settlements. The control strategy is similar to that invoked in Table 1. The general message from the table is that Roman road density is statistically strongly correlated with early economic activity, featuring elasticities between 0.6 and 1.1; and that roads are significant at the five percent-level or better, regardless of which fixed effects or controls are added.

Table 2

The controls themselves appear to enter in a meaningful way. Briefly, the negative estimate for latitude (column 1) indicates that the density of major settlements by the fifth century CE declines as one moves from the southern parts of the empire and to the north, probably testifying to the importance of the Mediterranean basin during antiquity. Further, we find more major settlements at low levels of elevation and in areas with productive agriculture (column 4); close to a coast and a natural harbour (column 5).

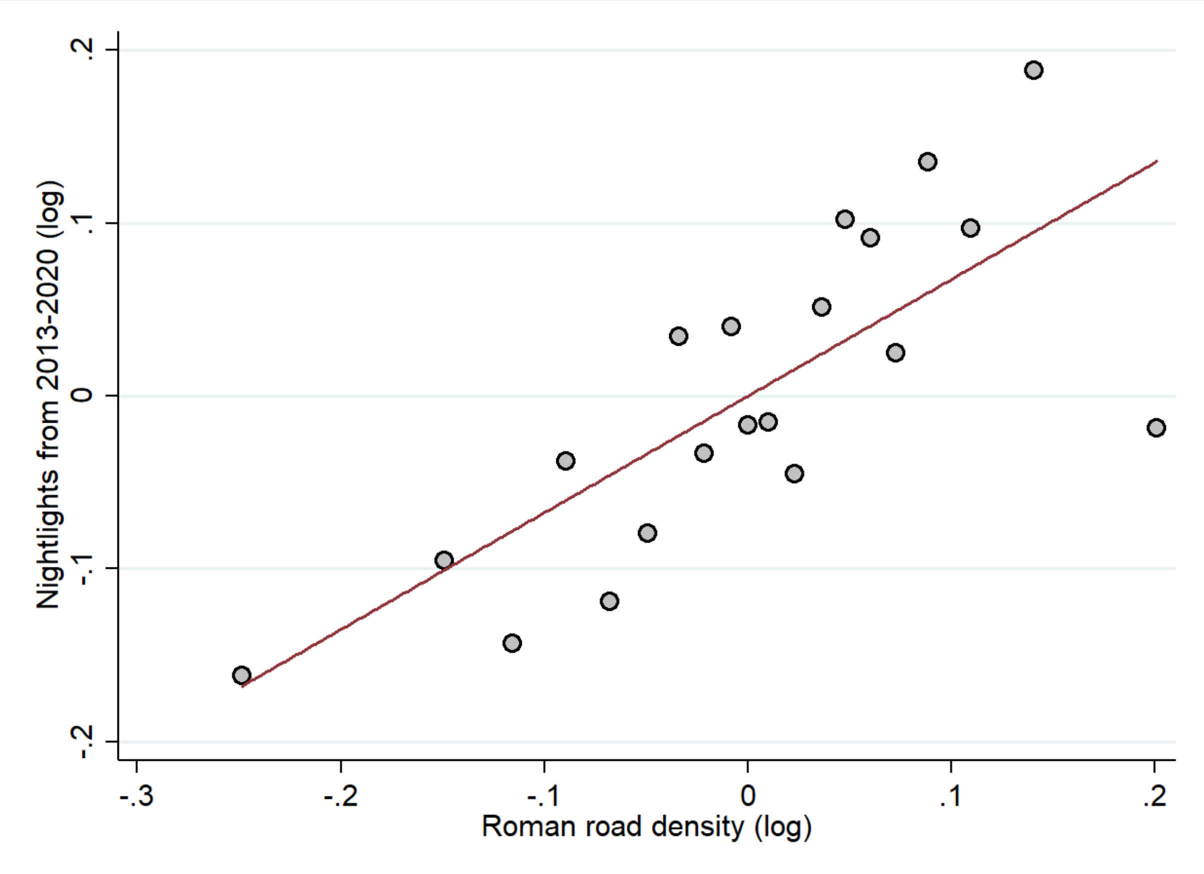
In the two subsequent tables, 3 and 4, we shift focus to contemporaneous economic activity, measured by nightlights (Table 3) and population density (Table 4).

Table 3, Table 4

Once again, we find a statistically strong signal from ancient roads on current economic activity. Regardless of controls, or exact choice of measure of economic activity, the estimate for Roman road density is significant in almost all columns at the one percent level or better. Figure 6 and 7 depict the binned added variables plot for contemporaneous nightlights and population density, respectively, in the context of our full

[26] After partialing out controls, we divide the sample into 20 equal sized bins, average the residualized Roman road density and modern road density within these bins, and plot the results in a scatter diagram.

Figure 6: Conditional relationship between average nightlight intensity in 2013-2020 and Roman road density in 117 CE within the former Roman Empire



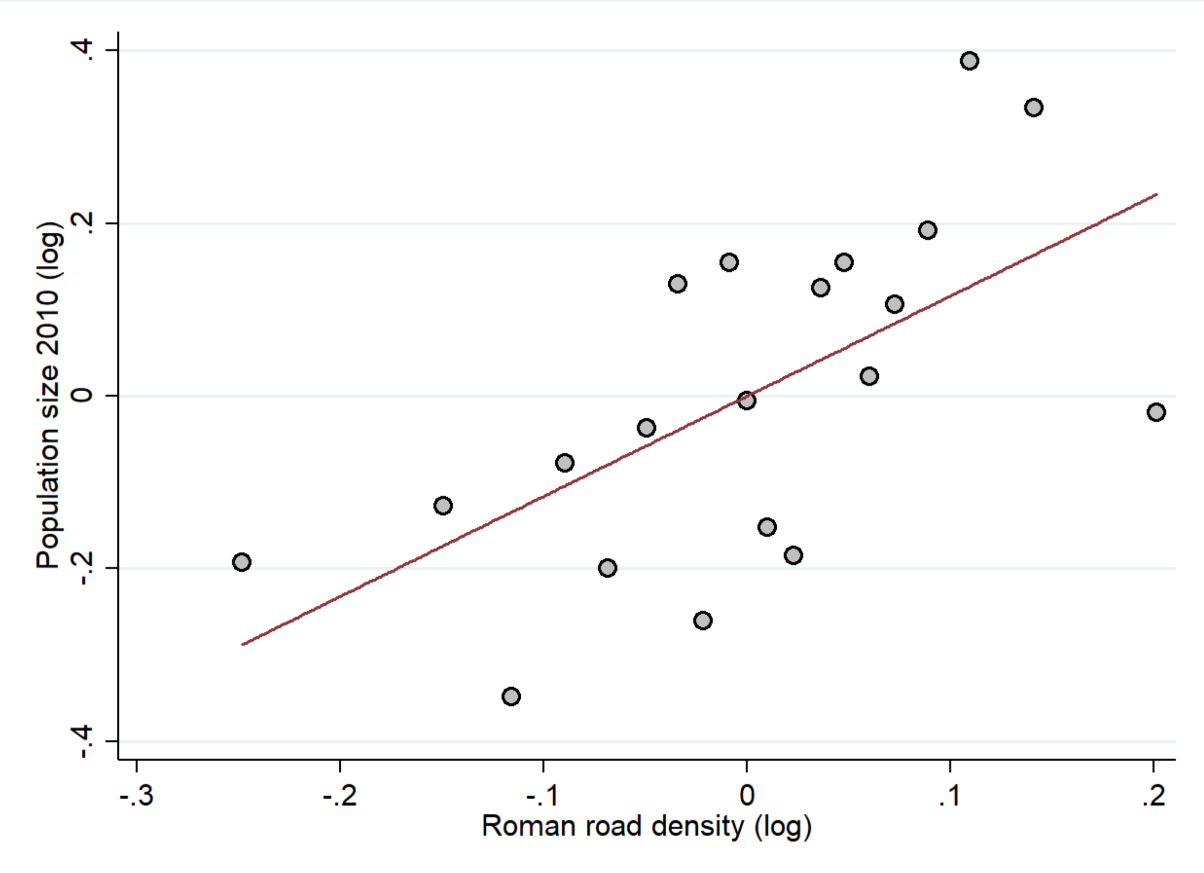
The figure shows the conditional binned residual scatter plot of the relationship between average nightlight intensity (in logs) in 2013-2020 and Roman road density (in logs) in 117 CE for 693 grid cells within the former Roman empire. The binned scatter plot groups the x-axis variable into equal-sized bins, computes the mean of the x-axis and y-axis variables within each bin, then creates a scatter plot of these data points. The underlying regression in Table 3, column 7 controls for the full set of geographical controls, as well as for country and language fixed effects, and hence estimate the within-country impact of historical Roman settlements on contemporaneous population levels. See text and Appendix for exact variable definitions.

specification (Column 7 in the two tables). The strong positive (partial) correlations do not appear to be driven by outliers.

The point estimate for the controls are broadly consistent with priors. In both tables we find that economic activity tends to decline as one moves away from the coast. Also, economic activity tends to decline at higher levels of elevation, and with distance to current country capitals. These commonalities are consistent with the notion that both measures are reasonable proxies for economic activity.

At the same time, the result do not always line up. When we examine the determinants of nightlights we find a positive latitude gradient, but this is not the case when

Figure 7: Conditional relationship between population in 2010 and Roman road density in 117 CE within the former Roman Empire



Note: The figure shows the conditional binned residual scatter plot of the relationship between population level (in logs) in 2010 and Roman road density (in logs) in 117 CE for 693 grid cells within the former Roman empire. The binned scatter plot groups the x-axis variable into equal-sized bins, computes the mean of the x-axis and y-axis variables within each bin, then creates a scatter plot of these data points. The underlying regression in Table 4, column 7 controls for the full set of geographical controls, as well as for country and language fixed effects, and hence estimate the within-country impact of historical Roman settlements on contemporaneous population levels. See text and Appendix for exact variable definitions.

population density is used. On the other hand, the post-1500 potential supply of calories seem to matter to population density, but does not help to explain the variation in nightlights. A potential explanation could be, that there need not be a perfect match between where people live and where they work. To see what this implies, suppose population density captures place of residency to a relatively greater extent than nightlights. Then the positive correlation with caloric suitability could be due to the fact that cities historically usually were located near rich agrarian hinterlands (Henderson *et al.*, 2016). The location of, say, a factory is potentially less path dependent than a city. Indeed, in recent times production has moved out of city centers to capitalize on lower land prices. Under a similar logic, latitude apparently influences productivity on-the-job more than the location of population centers.

Overall, Roman road density appears strongly associated with economic activity, both in the past and in the present. In almost every specification, statistical significance at the 1 percent level is attained, and the economic significance is quite substantial. In our full specifications, we find that urban economic activity during antiquity rises by about 0.5 percent for every percentage point increase in road density. For modern levels of population density, we find elasticities in the range 1.1 - 3, depending on the specification.

Importantly, in the Appendix, we examine the robustness of the link between ancient roads and all our outcome variables. In particular, examine whether the results are affected by decreasing the size of grid cells (Tables A4-A5) and if we refrain our analysis to a sample of only certain roads (rather than both certain and uncertain roads) (Tables A6-A7). Overall the results reported in tables carry over all these robustness checks, apart from in Table A6, column 2, where the positive estimate for Roman roads on settlements is insignificant when only certain roads are used.

4.3 Exploring channels I: Persistence and non-persistence

A key question regarding the results above is whether they reflect a causal impact of ancient roads on modern roads, and ultimately economic activity today. Naturally, the Roman roads are strongly predetermined, so reverse causality is not a concern. But it seems hard to rule out that underlying structural characteristics, perhaps notably of a geographic nature, could be driving both the intensity of Roman road treatment and the outcomes in focus. That is, despite our best efforts, the results may suffer from omitted variable bias.

In the present section we explore the likelihood that our results can be accounted for in this manner, by exploiting the remarkable abandonment of wheeled transport in the Middle East and North Africa (MENA) during the second half of the first millennium CE (Bulliet, 1990 [1975]). This event is an astonishing fact of world history. Perhaps especially since wheeled transport has had a very long history in the Middle East before its abandonment. The first instances of primitive two-wheeled carts, drawn by oxen or later by horses were found already in the earliest civilizations of ancient Mesopotamia, for example. Such transportation was clearly facilitated by roads. As mentioned above, notable roads were built in Persia during the Achaemenid period around 500 BCE. But during the Roman era the roads became more pervasive and better constructed. This frames the puzzle: why did wheeled transport decline and disappear under those circumstances?

4.3.1 Empirical strategy: The regional loss of wheeled transport

Bulliet (1990 [1975]) argues that the key proximate reason for the abandonment of the wheel was the emergence of the camel caravan as a more cost effective mode of transport of goods in the region. Camels outperform horses and oxen when it comes to stamina in desert-type conditions. The cost advantage during antiquity can be supported by data from Diocletian's price edict in 301 CE, which suggests a roughly 20 percent cost advantage in transport of goods by way of camel, relative to oxen.[27] To an economist, this seems like a very reasonable explanation. But it immediately prompts the question of why the ox-carriage then continued to dominate land-based transport until the first half of the first millenium CE? After all, the domestication of the camel on the Arabian Peninsula pre-dates the Roman era by millenia (Almathen *et al.*, 2016).

Bulliet's argument is that a series of developments had to take place before the camel could emerge as the dominant mode of transport in MENA. First, the emergence of a new type of camel saddle by 100 BCE made it possible for camel herding tribesmen

[27] The objective of the edict was to stabilize prices in the region, which makes it probable that the relative prices, stipulated by the edict, were based on relative cost differences. From the edict it can be calculated that the price of transporting a given amount of goods (in Roman pounds) over a given distance was 20 percent higher per oxen than per camel. See Bulliet (1990, Ch. 1) for further discussion.

to utilize new types of effective weapons. This improved the military strength of ethnic groups that centuries earlier had perfected camel breeding, which allowed them to gradually gain control of the trade routes and, as a consequence, gain political power as well. Second, another important factor was the decline of Rome and the ultimate rise of Islam as a key power factor in the Mediterranean. As also forcefully argued by Henri Pirenne (2012 [1937]), the ensuring decline in long distance trade across the Mediterranean allowed for an increasing importance of inland trade routes within the former Roman empire, which, in the case of the MENA supported caravan transport. However, horse or ox-drawn carts remained the main mode of inland transport in Europe. Therefore it is not surprising that while the Roman roads continued to be maintained and in use in Europe (Glick, 2005 [1979}; Hitchener, 2012), where wheeled vehicles dominated land-based transport, the same does not seem to be have been the case in the remaining regions of the Roman empire where the caravan took over (Bulliet, 1990).[28]

The implication of these developments is that since ancient roads fall into disrepair in the MENA region, to a much greater extent than in Europe, one should expect to see much less persistence in infrastructure density. The argument is simply that more than a millennium of disrepair most likely would erase the traces of the ancient infrastructure to a considerable extent, and when the importance of maintaining or building roads reappears in North Africa and the Middle East – with the advent of the automobile – the principles underlying road planning almost certainly *differed* from those that directed the planning of the Roman roads. In Europe, where the ancient roads persisted to a greater extent, modern roads are more likely to be built in place of the ancient roads. As a result, it would seem highly unlikely that modern road density would line up with ancient road density in the Middle East and North Africa whereas persistence would be more likely a priori in Europe.

The potentially differentiated degree of persistence in road density across regions of the Empire holds stark implications for the influence of Roman roads on comparative development: one should expect an influence from Roman road density on economic

[28]It is worth noting that, within Europe, the Roman roads left a mark on the landscape even in the places where they fell into disrepair thus supporting persistence. As observed by McGrath (2016, p. 15):"*But Roman roads proved persistent. Many paved and cobbled roads remained passable. And while the surface of metalled roads - the majority of stretches - was soon compromised by usage, frost and encroaching vegetation, the road beds continued to serve as transport conduits.*"

activity today only where persistence in infrastructure is found. Hence Roman roads should be of little importance to contemporary comparative development within the MENA region, while holding explanatory power within Europe. At the same time, one should expect a positive influence of Roman roads on economic activity in *all* regions during antiquity, *before* the abandonment of the wheel in North Africa and the Middle East.

These considerations lead to a straightforward test. We re-estimate equation (1) on subsamples: Europe and MENA, respectively. In this setting we expect to see persistence of an influence of Roman roads only within the European part of the Empire. This testing strategy allows us to assess the likelihood that our results above are driven by omitted variable bias. If indeed Roman roads do not predict modern road density in the MENA region, there is little reason why Roman roads should hold explanatory power vis-a-vis contemporary comparative development. If a significant link between past infrastructure investments and current economic activity arises in spite of this, the link is likely spurious or driven by unobserved geographic determinants of roads and economic development. Naturally, one might imagine that Roman roads could influence long run development through some type of cultural or institutional channel. But that concern is virtually eliminated in light of our extensive controls for current institutions and cultural variation, through country fixed effects and language group fixed effects in all regressions. Accordingly, the abandonment of the wheel experiment in effect allows us to explore the channel through which our baseline results come about.

Before we turn to the results one further issue is worth raising. Today the MENA region is considerably poorer, on average, than Europe. Perhaps the factors that stifled economic development in this region would also serve to mollify the explanatory power of the past? That is, perhaps an absence of a "signal" from the past, in this region, would have little to do with the mechanism in focus: that the abandonment of the wheel diminished the persistence of infrastructure and therefore diminished the persistence in comparative economic development. The key aspect to notice, however, is that this concern involves comparative development *across* regions. The strategy employed below involves looking *within* regions; regions that ultimately followed separate development trajectories overall. While the persistence of infrastructure in Europe *relative* to, say, the Middle East, may have many causes, the tests conducted below involve asking if infrastructure was persistent within the MENA region and, by extension, whether ancient infrastructure predicts comparative development within the MENA.

4.3.2 Empirical results

In Table 5 we begin by examining the correlation between Roman road density and our measure of economic activity by the end of antiquity in columns 1-2. The control set is the one in our full specification (cf column 7 in Tables 1-4).

Table 5

As is evident, across grid cells within Europe and within MENA, respectively, there is more economic activity in places with greater density of Roman roads. A natural interpretation of these findings is that by the end of antiquity areas more connected to the Roman road network benefited on net terms. Hence, prior to the abandonment of wheeled vehicles there is a positive influence from roads on comparative development, regardless of which region of the empire we focus on. If anything, the economic significance of the link appears stronger in MENA than within Europe.

If we then turn attention to contemporary outcomes, results change markedly. As seen from columns 3-4, Roman road density holds statistically significant predictive power within Europe, with respect to modern-day road density, whereas the same is not true for MENA. In fact, the estimate in column 4 is negative and not significant. The non-significance is consistent with what one would expect in the aftermath of the abandonment of the wheel-experiment. As ancient roads are left to decay they ultimately become a less reliable predictor of modern road location in the MENA. We will briefly come back to this issue below.

In the remaining columns 5-8 we turn attention to modern day economic activity. It is evident that whereas Roman roads hold strong predictive power over comparative development within Europe, the statistical significance levels are dramatically smaller within the MENA sample. In light of the absence of persistence in road density, these results are revealing, strongly suggesting that the explanatory power of Roman roads on current economic development is driven by the persistence of the road network.

In Table 6, we probe deeper into the *non-persistence* of Roman roads in the MENA region, and construct a proxy for the location of caravan routes for transportation, using digitized maps from recent research by Lado *et al.* (2020), who in turn rely on Bulliet (1990). Analogous to how we constructed the Roman road variable, we measure the share of a grid cell covered by a buffer zone of 5 km on either side of a caravan route. We then analyze whether the presence of Roman roads is in some way correlated with the caravan route density, restricting our sample to the grid cells in the MENA that had at least some portion of Roman roads during antiquity.[29]

Table 6

The striking result in Table 6 is that the estimates for Roman roads are small but consistently negative and at a fairly stable level, and statistically indistinguishable from zero in the richest specifications. The statistical insignificance is reassuring by itself: later caravan trade routes in MENA appear as orthogonal to the former road-based routes, when the caravan trade routes flourished with the expansion of Islam after 632 CE. However, the consistently negative (and in some cases significant) correlation between Roman roads and caravan resonates strongly with Pirenne (2012 [1937]), who argued that the rise of Islam led to the fall of the Mediterranean from a trade perspective, and moved away economic activity from it, thereby stimulating development in more northern parts of Europe, but also from a MENA perspective: It meant that activity moved away from centers of economic and political influence close to the Mediterranean, and instead towards the (inland) new centers in the MENA region. This tendency might potentially also be thought of as a reconfiguration of the transportation infrastructure in the eastern provinces in a manner reminiscent of the reorientation of the urban network in Roman Britain after the collapse of Roman institutions there, as described by Michaels and Rauch (2017). We believe that this finding is interesting by itself, and a potential area for further future research too.

4.4 Exploring Channels II: Market towns

The previous section showed that there was a persistence in economic activity since Roman times in Europe, where wheeled transport remained a central part of the transport system. In particular, we demonstrated in Figure 5 that the Roman road network

[29]We transform the caravan percentage measure by taking log(1+share of caravan buffer in cell), just as with the Roman road measure.

explained a great deal of the variation in contemporary European road infrastructure. But what were the channels through which Roman road investments had a positive impact on modern road infrastructure over a 1.5 millennia? In particular, how could Roman public infrastructure investments have a lingering impact even after the collapse of Western Rome in 476 CE, when political disorder and unstable state formations replaced Roman rule for several centuries?

We hypothesize that one key mechanism for the persistence of economic activity was the emergence of market towns from the early medieval period to the modern era.[30] Roman areas with a substantial level of historical investments in public infrastructure should simply be able to offer more suitable and more accessible locations for market places where people could meet and exchange goods. These market places might, in turn, later be connected with modern roads during the twentieth century.

In order to test this hypothesis, we use data from Cantoni and Yuchtman (2014) on the emergence of 2256 market towns in Germany from 850 CE onwards. The data was extracted from the *Deutsches Städtebuch* and captures information about when a market privilege was first granted to a town by the emperor or a local lord.[31] We combine this data with the Border Regression Discontinuity (BRD) design of Wahl (2017), comparing outcomes within a 100 km corridor on both sides of the historical Roman *limes* in contemporary Germany. Wahl (2017) demonstrates that there was a very high intensity of Roman road construction within this corridor and that a large fraction of the historical roads are currently located very close to pixel cells with modern highways. As in our estimations above, the unit of analysis in Wahl (2017) is a pixel cell but with a very small size; only 0.86 square kilometers. The corridor along the *limes* contains 181,915 such cells, within which 398 towns are located that were granted market privileges, most of them during the 12-16th centuries.

The main empirical equation that we estimate is

$$Market_{gs} = \alpha + \beta \cdot Roman_{gs} + f(\mathbf{Z}_{gs}) + \mathbf{X}'_{gs}\mathbf{fl} + \delta_s + \epsilon_{gs} \tag{2}$$

[30]See Michaels and Rauch (2017) for an analysis of the persistence of the Roman urban network in France vs England.

[31]Receiving a market right typically also gave the local lord the right to tax trade in that market, coin money, and require its use in the market. Lords were, in turn, responsible for the protection of merchants, the provision of courts to settle disputes, and the establishment of standard measures (Cantoni and Yuchtman, 2014, p 832).

In this expression, the outcome variable is a dummy variable $Market_{gs}$ equal to 1 if there was a market town at any time in pixel cell g along border segment s. The main variable of interest is $Roman_{gs}$ which is a binary dummy for the Roman side of the *limes*. $f(\mathbf{Z}_{gs})$ is a flexible polynomial function of distances to border and polynomial functions of the latitude and longitude coordinates of each pixel cell g, following the standard in the BRD literature. $\mathbf{X}_{gs}$ is a vector of control variables (distance to the closest river, elevation, ruggedness and agricultural suitability). δ_s is a dummy for segments of the border corridor (there are five in total) and ϵ_{gs} is an error term.

The results are displayed in Table 7. The RD-specifications in columns (1)-(5) exactly mimic those of Wahl (2017, Table 3, columns 2-6), except that our outcome variable is a market town dummy rather than a measure of night lights intensity.

In all specifications, the Roman area dummy is positive and significant at the 0.05-0.1-level. In column (5), we include both latitude-longitude and distance-to-border polynomials, segment dummies and controls. The estimate .003 implies that market towns were 0.3 percent more likely to emerge in an area of size 0.86 km^2 in the historical Roman part of modern Germany relative to outside. Since the fraction of cells with market towns in our sample is 0.22 percent, this is an economically quite large effect.

Table 7

So far in Table 7, we have not employed any indicator for Roman road presence, as in the rest of our analysis. In columns (6)-(7), we focus on the Roman side in Wahl's German sample with 83,211 grid cells and use a dummy for whether the grid cell contains a Roman road in (6), and (log) distance to a Roman road in (7). The results are as predicted: The presence of a Roman road increases the probability of observing a later medieval market town with 0.3 percent, whereas distance to a Roman road has a negative influence on later market towns.

Overall, we believe that these results provide interesting perspectives on the roots of comparative development. While previous research has demonstrated how the observed persistence in economic development can arise due to variations in geographic initial conditions, either directly or indirectly via cultural or institutional change, our results draw attention to an important role for shocks with persistent influence. From the point of view of any given geographical sub-region, the emergence of the Roman

Empire, and with it the Roman road network, is best viewed as external. The modest importance of geography in dictating the location of roads (cf Section 3) illustrates that *second nature* processes can, to some extent independently of geography (or *first nature* processes), have a substantial impact on long-run comparative development. In our context in this paper, the persistence of the shocks, and thereby of comparative development, arise through a remarkable degree of persistence in road density across several millennia, in regions where the roads were deemed economically useful. Evidently, persistence in transport infrastructure investments is a potential source of persistence in comparative development.

5 Concluding remarks

The existing literature on comparative development has drawn attention to a remarkable pattern of persistence in economic activity: places featuring comparatively high levels of economic development long before the industrial revolution often seem to feature high levels of comparative development today. In the present project we examine the persistence of an important proximate source of economic activity: Public infrastructure investments.

Our analysis reveals that, within regions that used to be part of the Roman empire, infrastructure density is highly persistent. That is, Roman road density is generally a strong predictor of modern day road density. Moreover, Roman road density is generally a predictor of contemporary economic activity. These results are statistically strong and resilient to extensive controls including for contemporary institutions and cultural values. Taken at face value, these results suggest that infrastructure may be one important channel through which the persistence in comparative development comes about.

In examining whether our core results, linking early infrastructure to current-day infrastructure and economic activity, are likely to reflect causal relationships, we examine the remarkable historical case of the abandonment of the wheel that occurred in the Middle-East and North Africa (MENA) during the second half of the first millennium CE. We find that in the MENA region, Roman roads lose predictive power vis-a-vis modern day roads. Moreover, Roman road density does not predict current day economic activity within the MENA region. In contrast, in the European region, where the roads were maintained, our baseline results carry over.

The results from this natural experiment suggest quite strongly that our reduced form findings, linking Roman road density to current comparative development, are indeed due to the persistence of that type of infrastructure. Exploring further reasons behind this result, we find support for the idea that the emergence of market towns is a robust and economically meaningful mechanism to explain the persistence and the persistent effects of infrastructure over a remarkable period of 2000 years.

References

[1] Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). "When Should You Adjust Standard Errors for Clustering?" (No. w24003). National Bureau of Economic Research.

[2] Almathen, F., Charruau, P., Mohandesan, E., Mwacharo, J. M., Orozco-terWengel, P., Pitt, D., ... & Magee, P. (2016). "Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary" *Proceedings of the National Academy of Sciences*, 113(24), 6707-6712.

[3] Andersen, T. B., Dalgaard, C. J., & Selaya, P. (2016). "Climate and the emergence of global income differences" *Review of Economic Studies*, 83(4), 1334-1363.

[4] Angelucci, C., Meraglia, S., & Voigtländer, N. (2017). "The Medieval Roots of Inclusive Institutions: From the Norman Conquest of England to the Great Reform Act" (Working Paper No 23606). National Bureau of Economic Research.

[5] Ashraf, Q., & Galor, O. (2013). "The Out of Africa hypothesis, human genetic diversity, and comparative economic development" *American Economic Review*, 103(1), 1-46.

[6] Ashraf, Q. and O. Galor (2018). "The Macrogenoeconomics of Comparative Development", *Journal of Economic Literature* 56(3), 1119-1155.

[7] Barjamovic, G., T. Chaney, K. Cosar, and A. Hortacsu (2019) "Trade, merchants, and the lost cities of the Bronze Age" *Quarterly Journal of Economics,* 134(3) August 2019, 1455–1503.

[8] Banerjee, A., E. Duflo and N. Qian (2012) "On the Road: Access to Transportation Infrastructure and Economic Growth in China" mimeo.

[9] Benedictis, L, V. Licio, and A.M. Pinna (2018) "The long-term effect of the historical Roman road network: Trade costs of Italian provinces" mimeo.

[10] Bird, J. and S. Straub (2014) "The Brasilia Experiment: Road Access and the Spatial Pattern of Long-term Local Development in Brazil" *TSE Working Paper*, n. 14-495, May 2014

[11] Bosker, M., & Buringh, E. (2017). "City seeds: Geography and the origins of the European city system" *Journal of Urban Economics*, 98, 139–157

[12] Bosker, M., E. Buringh, and J.L. van Zanden (2013) "From Baghdad to London: Unraveling Urban Development in Europe, The Middle East and North Africa 800-1800" *Review of Economics and Statistics* 95(4), 1418-1437.

[13] Bowman, A. and A. Wilson (2009) *Quantifying the Roman Economy: Methods and Problems*. Oxford University Press.

[14] Bulliet, R. W. (1990 [1975]). *The Camel and the wheel*. Columbia University Press. Morningside Edition.

[15] Cantoni, D. and N. Yuchtman (2014) ”Medieval universities, legal institutions, and the commercial revolution” *Quarterly Journal of Economics*, 823-887.

[16] Chanda, A., Cook, C. J., & Putterman, L. (2014). ”Persistence of fortune: Accounting for population movements, there was no post-Columbian reversal” *American Economic Journal: Macroeconomics*, 6(3), 1-28.

[17] Chen, T., J. Kung and C. Ma (2017) ”Long Live Keju! The Persistent Effects of China’s Imperial Examination System”. Mimeo

[18] Chaves, I., S. Engerman, S., and J. Robinson (2014). ”Reinventing the Wheel: The Economic Benefits of Wheeled Transportation in Early Colonial British West Africa” In E. Akyeampong, R. Bates, N. Nunn, and J. Robinson (eds.), *Africa’s Development in Historical Perspective* (pp. 321-365). Cambridge: Cambridge University Press.

[19] Colella, F., R. Lalive, and M. Thoenig (2020) ”Inference with Arbitrary Clustering” IZA Discussion Paper n. 12584.

[20] Comin, D., Easterly, W., & Gong, E. (2010). Was the wealth of nations determined in 1000 BC?. *American Economic Journal: Macroeconomics*, 2(3), 65-97.

[21] Conley, T. (1999) ”GMM Estimations with Cross-Sectional Dependence” *Journal of Econometrics* 92, 1-45.

[22] Davies, H.E. (1998) ”Designing Roman roads” *Britannia*, 29: 1-16.

[23] Donaldson, D. (2018) ”Railroads of the Raj: Estimating the Impact of Transport Infrastructure” *American Economic Review* 108(4-5), 899-934.

[24] Fernald, J.G. (1999) ”Roads to Prosperity? Assessing the Link Between Public Capital and Productivity” *American Economic Review* 89(3), 619-638.

[25] Finley, M. (1973) *The Ancient Economy*, Berkeley: University of California Press.

[26] Fluckinger, M., E. Hornung, M. Larch, M. Ludwig, and A. Mees (2019) ”Roman transport network connectivity and economic integration” *Review of Economic Studies*.

[27] Franco, A.P., S. Galiani, and P. Lavado (2021) ”Long-term effects of the Inca Road” NBER Working Paper 28979

[28] Gabriel, R.A. (2002) *The Great Armies of Antiquity*, Westport: Greenwood.

[29] Galor, O. and Ö. Özak. ”The agricultural origins of time preference.” *American Economic Review* 106.10 (2016): 3064-3103.

[30] Glaeser, E. L., & Shleifer, A. (2002). Legal origins. *Quarterly Journal of Economics*, 117(4), 1193-1229.

[31] Glick (2005 [1979]). *Islam and Christian Spain in the early Middle Ages*. Brill.

[32] Gorodnichenko, Y., & Roland, G. (2017). ”Culture, institutions, and the wealth of nations” *Review of Economics and Statistics*, 99(3), 402-416.

[33] Gros, Pierre (1991) *La France gallo-romaine*, Paris: Nathan.

[34] Jedwab, R. and A. Moradi (2016) "The Permanent Effects of Transport Revolutions in Poor Countries: Evidence from Africa" *Review of Economics and Statistics* 98(2), 268-284.

[35] Henderson, V., A. Storeygaard and D. Weil (2012) "Measuring economifc growth from outer space" *American Economic Review* 102(2), 994-1028.

[36] Hornung, E. (2015) "Railroads and Growth in Prussia" *Journal of the European Economic Association* 13(4), 699-736.

[37] Hitchener, R.B. (2012). 'Roads, integration, connectivity and economic performance in the Roman Empire',in (S.E. Alcock, J. Bodel and R.J.A. Talbert, eds.), Highways, Byways, and Road Systems in the Pre-ModernWorld, pp. 222–30, Malden, MA: Wiley-Blackwell.

[38] Jedwab, R., & Moradi, A. (2016). "The permanent effects of transportation revolutions in poor countries: evidence from Africa" *Review of Economics and Statistics*, 98(2), 268-284.

[39] Howard, P. (2006) *The Fall of the Roman Empire: A New History of Rome and the Barbarians*, Oxford University Press.

[40] Huilery, E. (2009) "History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa" *American Economic Journal: Applied* 1(2), 176-215.

[41] Kelly, M. (2021) "Persistence, Randomization, and Spatial Noise" *UCD Centre for Economic Research Working Paper Series* WP2021/24.

[42] Krugman, P. (1991) "History vs Expectations" *Quarterly Journal of Economics* 106(2), 651-667.

[43] Lado, S. *et al.* (2020). "Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes" *Communications Biology*, doi.org/10.1038/s42003-020-1098-7.

[44] Landes, D (1998). *The Wealth and Poverty of Nations*. Norton and Company: New York.

[45] Langas, S. (1995), "Completeness of the Digital Chart of the World (DCW) database". UNEP/GRID-Arendal, Project No. 205026.

[46] Laurence, R. (1999) *The Roads of Roman Italy: Mobility and Cultural Change*, Routledge: London.

[47] Maloney, W.F., and F. Valencia Caicedo (2016) "The Persistence of (Subnational) Fortune" *Economic Journal* 126(598), 2363-2401.

[48] McCormick, M., G. Huang, G. Zambotti, and J. Lavash (2013), "Roman Road Network (version 2008)", DARMC Scholarly Data Series, Data Contribution Series #2013-5. DARMC, Center for Geographic Analysis, Harvard University.

[49] McGrath, Patrick (2016). ”Roman roads in Scotland: A contribution to measuring the long-run impact of transport infrastructure on local economic geography” MSc Thesis (University of Oxford).

[50] Michaels, G. (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.

[51] Michaels, G. and F. Rauch (2017) ”Resetting the Urban Network: 117-2012” *Economic Journal* 128(608), 378-412.

[52] Michalopoulos, S., & Papaioannou, E. (2013). ”Pre-Colonial ethnic institutions and contemporary african development” *Econometrica*, 81(1), 113-152.

[53] Mokyr, J. (1990) *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford: Oxford University Press.

[54] New, M., D. Lister, M. Hulme and I. Makin (2002). ”A high-resolution data set of surface climate over global land areas” *Climate research*, 21(1), 1-25.

[55] Nunn N. ”Historical Development”. In: Aghion P, Durlauf S Handbook of Economic Growth. Vol. 2. North-Holland ; 2014. pp. 347-402.

[56] Nunn, N., and D. Puga (2012). ”Ruggedness: The blessing of bad geography in Africa” *Review of Economics and Statistics*, 94(1), 20-36.

[57] Olsson, O. and D.A Hibbs, Jr (2005). ”Biogeography and Long-run Development” *European Economic Review* 49(4), 909-938.

[58] Pinhasi, R., J. Fort, and A. Ammerman (2005) ”Tracing the Origins and Spread of Agriculture in Europe” *PLOS Biology* 3(12), 2220-2228.

[59] Pirenne, H. (2012 {1937]) *Economic and Social History of Medieval Europe*, Harcourt: Brace.

[60] Ramankutty, N., J. A. Foley, J. Norman and K. McSweeney (2002). ”The global distribution of cultivable lands: current patterns and sensitivity to possible climate change” *Global Ecology and Biogeography*, 11(5), 377-392.

[61] Rappaport J. and J. Sachs (2003) ”The US as a coastal nation” *Journal of Economic Growth* 8, 5-46.

[62] Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development?. *Journal of Economic Literature*, 51(2), 325-369.

[63] Tabellini, G. (2010) ”Culture and Institutions: Economic Development in the Regions of Europe” *Journal of the European Economic Association* 8(4), 677-716.

[64] Talbert, R.J. (2000) *Barrington Atlas of the Greek and Roman World*, Princeton University Press.

[65] Temin (2012) *The Roman Market Economy*, Princeton University Press.

[66] Wahl, F. (2017) ”Does European Development Have Roman Roots? Evidence from the German Limes” *Journal of Economic Growth*, 22(3), pp. 313-349.

[67] Woolf, G. (1998) *Becoming Roman: The Origins of Provincial Civilization in Gaul.* Cambridge: Cambridge University Press.

[68] World GeoDatasets (2017), "Seamless Digital Chart of the World Base Map version 10.5". Available at http://www.worldgeodatasets.com/files/6214/9305/6335/WGDS_SDCW.pdf

[69] Åhlfeldt, J. (2017) *Digital Atlas of the Roman Empire* . Web publication available at http://dare.ht.lu.se/. Lund University, Lund.

Tables

Table 1: Roman Roads and Modern Roads

	Dependent variable: Modern roads						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln (1+) Roman roads	0.331***	0.240***	0.222***	0.115*	0.203***	0.214***	0.119*
	(0.055)	(0.058)	(0.062)	(0.063)	(0.062)	(0.063)	(0.065)
	[0.067]	[0.066]	[0.077]	[0.077]	[0.075]	[0.075]	[0.075]
ln Area	-0.003	0.006	0.005	0.009	0.008	0.005	0.008
	(0.006)	(0.007)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
ln Latitude	0.322***						
	(0.043)						
ln (10+) Longitude	-0.013						
	(0.039)						
ln Longitude squared	0.002						
	(0.008)						
ln (50+) Elevation				-0.051***			-0.049***
				(0.011)			(0.012)
ln Post-1500 caloric suitability				0.012**			0.011**
				(0.005)			(0.005)
ln Distance to coast					-0.010		0.001
					(0.008)		(0.008)
ln Distance to major river					-0.006		0.002
					(0.006)		(0.006)
ln Distance to natural harbor					-0.011**		-0.012**
					(0.005)		(0.006)
ln Distance to Roman Empire border						-0.007	-0.004
						(0.013)	(0.013)
ln (1+) Number of mines						-0.003	-0.001
						(0.008)	(0.008)
ln Distance to capital						-0.013	-0.009
						(0.010)	(0.010)
ln (1+) Distance to coal						0.003	-0.003
						(0.005)	(0.005)
ln Distance to university						-0.008	0.000
						(0.011)	(0.011)
ln Distance to Wittenberg						-0.026	-0.044
						(0.054)	(0.051)
Country FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Country-language FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	693	693	693	693	693	693	693
R^2	0.232	0.391	0.475	0.515	0.488	0.481	0.524

Notes: Roman roads and modern roads are defined as the log of one plus the fraction of a 5 km buffer around, respectively, the Roman and modern road system that lies wihin the total area of a countrycell. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 2: Roman Roads and Settlements

	Dependent variable: Roman Settlements in 500 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln (1+) Roman roads	1.096^{***}	0.811^{***}	0.832^{***}	0.577^{***}	0.728^{***}	0.778^{***}	0.499^{**}
	(0.169)	(0.178)	(0.189)	(0.203)	(0.195)	(0.187)	(0.205)
	[0.065]	[0.215]	[0.184]	[0.176]	[0.176]	[0.189]	[0.188]
ln Area	0.168^{***}	0.171^{***}	0.199^{***}	0.205^{***}	0.218^{***}	0.186^{***}	0.200^{***}
	(0.019)	(0.021)	(0.017)	(0.018)	(0.019)	(0.017)	(0.020)
ln Latitude	-0.813^{***}						
	(0.174)						
ln (10+) Longitude	-0.118						
	(0.126)						
ln Longitude squared	0.037						
	(0.027)						
ln (50+) Elevation				-0.107^{***}			-0.069
				(0.041)			(0.045)
ln Pre-1500 caloric suitability				0.037^{**}			0.039^{**}
				(0.018)			(0.017)
ln Distance to coast					-0.062^{***}		-0.039
					(0.023)		(0.026)
ln Distance to major river					-0.027		-0.018
					(0.026)		(0.027)
ln Distance to natural harbor					-0.053^{**}		-0.058^{**}
					(0.024)		(0.024)
ln Distance to Roman Empire border						0.014	-0.004
						(0.028)	(0.030)
ln (1+) Number of mines						0.115^{**}	0.123^{***}
						(0.046)	(0.046)
Country FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Country-language FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	693	693	693	693	693	693	693
R^2	0.189	0.324	0.394	0.407	0.416	0.401	0.430

Notes: Roman roads is log of one plus the fraction of a 5 km buffer around the Roman road system that lies wihin the total area of a country-cell. Roman settlements is log of one plus the number of major settlements within the country-cell in CE 500. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 3: Roman Roads and Nightlights

	Dependent variable: Nightlights from 2013-2020						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln (1+) Roman roads	1.509***	1.346***	1.199***	0.714***	1.037***	1.094***	0.675***
	(0.193)	(0.178)	(0.166)	(0.164)	(0.162)	(0.163)	(0.158)
	[0.213]	[0.208]	[0.200]	[0.157]	[0.171]	[0.187]	[0.123]
ln Area	-0.040	-0.011	-0.010	0.026	0.023	-0.025	0.029
	(0.029)	(0.025)	(0.023)	(0.023)	(0.023)	(0.028)	(0.026)
ln Latitude	-0.412**						
	(0.195)						
ln (10+) Longitude	-0.099						
	(0.214)						
ln Longitude squared	-0.010						
	(0.041)						
ln (50+) Elevation				-0.290***			-0.232***
				(0.034)			(0.038)
ln Post-1500 caloric suitability				0.017			-0.003
				(0.016)			(0.016)
ln Distance to coast					-0.101***		-0.067**
					(0.025)		(0.028)
ln Distance to major river					-0.041**		-0.001
					(0.021)		(0.020)
ln Distance to natural harbor					-0.046**		-0.043**
					(0.020)		(0.019)
ln Distance to Roman Empire border						-0.006	-0.017
						(0.042)	(0.041)
ln (1+) Number of mines						-0.004	0.006
						(0.031)	(0.029)
ln Distance to capital						-0.191***	-0.197***
						(0.047)	(0.048)
ln (1+) Distance to coal						-0.002	-0.030*
						(0.017)	(0.016)
ln Distance to university						-0.052	-0.018
						(0.043)	(0.041)
ln Distance to Wittenberg						-0.127	-0.302*
						(0.184)	(0.167)
Country FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Country-language FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	691	691	691	691	691	691	691
R^2	0.179	0.465	0.562	0.621	0.598	0.597	0.668

Notes: Roman roads is log of one plus the fraction of a 5 km buffer around the Roman road system that lies wihin the total area of a country-cell. Nightlights is log of the average light intensity measured at night by satelite from 2013-2020. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 4: Roman Roads and Population in 2010

	Dependent variable: Population in 2010						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln (1+) Roman roads	3.017***	2.638***	2.624***	1.215***	2.246***	2.464***	1.160***
	(0.414)	(0.426)	(0.396)	(0.345)	(0.375)	(0.376)	(0.326)
	[0.487]	[0.584]	[0.503]	[0.378]	[0.392]	[0.484]	[0.326]
ln Area	1.002***	1.029***	1.099***	1.115***	1.171***	1.060***	1.115***
	(0.064)	(0.069)	(0.039)	(0.039)	(0.037)	(0.040)	(0.038)
ln Latitude	2.292***						
	(0.507)						
ln (10+) Longitude	-0.395						
	(0.347)						
ln Longitude squared	0.055						
	(0.070)						
ln (50+) Elevation				-0.510***			-0.431***
				(0.076)			(0.082)
ln Post-1500 caloric suitability				0.246***			0.214***
				(0.056)			(0.053)
ln Distance to coast					-0.224***		-0.117**
					(0.055)		(0.054)
ln Distance to major river					-0.113**		-0.036
					(0.048)		(0.049)
ln Distance to natural harbor					-0.078*		-0.088**
					(0.046)		(0.040)
ln Distance to Roman Empire border						0.075	0.077
						(0.079)	(0.072)
ln (1+) Number of mines						-0.062	-0.046
						(0.061)	(0.055)
ln Distance to capital						-0.461***	-0.390***
						(0.079)	(0.075)
ln (1+) Distance to coal						-0.032	-0.094***
						(0.033)	(0.030)
ln Distance to university						-0.058	0.042
						(0.075)	(0.079)
ln Distance to Wittenberg						-0.483	-0.726**
						(0.324)	(0.316)
Country FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Country-language FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	693	693	693	693	693	693	693
R^2	0.593	0.701	0.769	0.812	0.784	0.787	0.829

Notes: Roman roads is log of one plus the fraction of a 5 km buffer around the Roman road system that lies wihin the total area of a country-cell. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 5: Sample Split: Roman Roads and Development in Europe and MENA

Dependent variable:	Settlements in 500 CE		Modern roads		Nightlights from 2010-2013		Population in 2010	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	Europe	MENA	Europe	MENA	Europe	MENA	Europe	MENA
ln (1+) Roman roads	0.470*	0.599*	0.208***	-0.115	0.783***	0.405	1.405***	0.717
	(0.255)	(0.358)	(0.072)	(0.091)	(0.191)	(0.282)	(0.390)	(0.630)
	[0.171]	[0.371]	[0.084]	[0.071]	[0.155]	[0.269]	[0.410]	[0.637]
ln Area	0.204***	0.235***	0.026***	-0.052***	0.071**	-0.069*	1.180***	1.045***
	(0.023)	(0.050)	(0.009)	(0.013)	(0.029)	(0.039)	(0.035)	(0.096)
ln (50+) Elevation	-0.090*	-0.060	-0.061***	-0.011	-0.226***	-0.300***	-0.456***	-0.587***
	(0.054)	(0.096)	(0.015)	(0.021)	(0.046)	(0.078)	(0.089)	(0.206)
ln Pre-1500 caloric	-0.012	0.051**						
suitability	(0.018)	(0.020)						
ln Distance to coast	-0.013	-0.095*	0.010	0.006	-0.081**	-0.021	-0.103*	-0.101
	(0.030)	(0.057)	(0.011)	(0.011)	(0.038)	(0.047)	(0.054)	(0.125)
ln Distance to major	-0.000	-0.034	0.010	0.005	0.034	-0.058	0.073	-0.306***
river	(0.027)	(0.068)	(0.007)	(0.010)	(0.023)	(0.042)	(0.048)	(0.111)
ln Distance to	-0.030	-0.071*	-0.014*	-0.018**	-0.028	-0.104***	-0.060	-0.146*
natural harbor	(0.031)	(0.040)	(0.009)	(0.007)	(0.027)	(0.030)	(0.051)	(0.081)
ln Distance to Roman	0.011	0.077	-0.008	-0.141	-0.042	-0.494	0.033	0.195
Empire border	(0.031)	(0.292)	(0.013)	(0.099)	(0.044)	(0.403)	(0.076)	(0.940)
ln (1+) Number of	0.071	0.394**	-0.008	0.027	-0.021	0.148	-0.095*	0.143
mines	(0.045)	(0.167)	(0.008)	(0.022)	(0.031)	(0.095)	(0.056)	(0.182)
ln Post-1500 caloric			0.011*	0.017***	-0.005	0.013	0.115***	0.240***
suitability			(0.006)	(0.006)	(0.023)	(0.019)	(0.039)	(0.080)
ln Distance to			-0.013	-0.003	-0.178***	-0.316***	-0.335***	-0.667***
capital			(0.012)	(0.014)	(0.063)	(0.068)	(0.099)	(0.162)
ln (1+) Distance to			-0.002	-0.002	-0.025	-0.046	-0.085***	-0.012
coal			(0.005)	(0.010)	(0.016)	(0.045)	(0.029)	(0.119)
ln Distance to			0.006	-0.168*	-0.035	0.610*	-0.011	1.041
university			(0.011)	(0.098)	(0.041)	(0.310)	(0.074)	(0.968)
ln Distance to			-0.073	0.671**	-0.470**	1.483	-0.862***	-0.996
Wittenberg			(0.055)	(0.280)	(0.188)	(1.047)	(0.329)	(2.360)
Country-language FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	468	225	468	225	466	225	468	225
R^2	0.428	0.409	0.519	0.562	0.700	0.677	0.896	0.729

Notes: Roman roads and modern roads are defined as the log of one plus the fraction of a 5 km buffer around, respectively, the Roman and modern road system that lies wihin the total area of a country-cell. Roman settlements is log of one plus the number of major settlements within the country-cell in 500 CE. Nightlights is log of the average light intensity measured at night by satelite from 2013-2020 CE. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 6: Caravan Routes and Roman Roads

	Dependent variable: Camel caravan routes						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln (1+) Roman roads	-0.028**	-0.039**	-0.030*	-0.029*	-0.026	-0.029	-0.029
	(0.012)	(0.019)	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)
	[0.012]	[0.017]	[0.015]	[0.016]	[0.014]	[0.015]	[0.015]
ln Area	0.003**	0.004**	0.006***	0.006**	0.003	0.006***	0.004
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
ln Latitude	-0.123***						
	(0.030)						
ln (10+) Longitude	-0.044*						
	(0.023)						
ln Longitude squared	0.009**						
	(0.004)						
ln (50+) Elevation				0.000			-0.004
				(0.004)			(0.007)
ln Pre-1500 caloric suitability				-0.000			0.000
				(0.001)			(0.002)
ln Distance to coast					0.003		0.004
					(0.002)		(0.003)
ln Distance to major river					0.001		0.002
					(0.002)		(0.002)
ln Distance to natural harbor					-0.000		-0.001
					(0.001)		(0.002)
ln Distance to Roman Empire border						0.005	0.008
						(0.004)	(0.009)
ln (1+) Number of mines						-0.005	-0.005
						(0.004)	(0.004)
Country FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Country-language FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	225	225	225	225	225	225	225
R^2	0.161	0.351	0.505	0.505	0.510	0.507	0.513

Notes: Roman roads and caravan routes are defined as the log of one plus the fraction of a 5 km buffer around, respectively, the Roman and caravan route system that lies wihin the total area of a countrycell. The analysis is performed on country-cells within the Roman empire in the Middle East and North Africa containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 7: Medieval markets and the Roman empire: RD and OLS regressions

	Dependent variable: Medieval market dummy						
	German border area					Roman part	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Roman area (dummy)	0.001**	0.004**	0.004**	0.003	0.003*		
	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)		
Roman road in pixel (dummy)						0.003*	
						(0.001)	
ln dist. to Roman road							-0.001**
							(0.000)
Order of coord. polyn.	3rd		3rd	3rd	3rd		
Order of dist. polyn.		8th	8th	8th	8th		
Segment dummies	No	No	No	Yes	Yes	No	No
Controls	No	No	No	No	Yes	Yes	Yes
Observations	181,915	181,915	181,915	181,915	181,915	83,211	83,211
R^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: This table documents the effect of the Roman Empire on market rights in the pre-modern period. The unit of analysis is 181,195 grid cells (size 0.86 km2) along the Roman limes from Wahl (2017) and the outcome variable is a dummy equal to 1 if there was a market town during the extended medieval period in the grid cell. The data on market towns was obtained from Cantoni and Yuchtman (2014). The right-hand side empirical specification is identical to that inWahl (2017, Table 3, columns 2-6). Heteroskedasticity robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 pct. level, ** at the 5 pct. level, and * at the 10 pct. level.

APPENDIX

Roman Roads to Prosperity: Persistence and Non-Persistence of Public Infrastructure

Carl-Johan Dalgaard*

Nicolai Kaarsen†

Ola Olsson‡

Pablo Selaya§

November 16, 2021

1 Additional data definitions and sources

Oppida were Celtic settlement sites that functioned as economic and political centres during the last two centuries BCE continuing to the first century CE before in Britain and Europe. The settlement area of oppidas was often hundreds of hectares, and they typically accommodated thousands of regular inhabitants. They were also characterized by being heavily fortified, and having a timber-laced stone-faced murus gallicus or a 'Gallic wall'.

We take data for the location of oppida during the La Tène CD period preceding the beginning of Roman expansion, from the Archaeology Data Service at the University of York.[1] Given that oppida were present only in Britain and continental Europe, we create an envelope of the buffer of 100 km around all oppida, as an estimate of the influece area of the La Tène culture. Within this area, similar to our procedure to estimate the area of influence of roads, we construct a buffer of 5 km

*University of Copenhagen, CAGE (Warwick) and CEPR (London). carl.johan.dalgaard@econ.ku.dk

†Danish Economic Councils. nik@dors.dk

‡University of Gothenburg. ola.olsson@economics.gu.se

§University of Copenhagen. pablo.selaya@econ.ku.dk

[1] The complete documentation can be found at http://archaeologydataservice.ac.uk/archiveDS/archiveDownload?t=arch-408-1/dissemination/zip/Atlas/GIS_data/documentation/ATLAS_METADATA.pdf.

around each oppidum, and compute the percentage of that area within each country-cell.[2]

Language areas are a complete set of language fixed effects, or indicator variables of ethnic languages recorded in the World Language Mapping System Database version 3.01 (WLMS, which is a dataset containing polygons for the linguistic homelands of more than 7,000 ethnic languages around the globe). We follow Andersen et al (2016) in the construction of these variables and, basically, we consider the predominant ethnic language to be the one with the largest area in cases where a country-cell has more than one ethnic language; and we assign a separate dummy variable that represents the excluded language category in each country, in country-cells where there are no specific ethnic languages recorded in WLMS.

Elevation is computed using the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010). This data set contains the average elevation in meters at the level of 30 arc-second cells. We aggregate elevation by averaging across smaller cells within each 1 by 1 degree cell.

Caloric suitability is computed by Galor and Özak (2016) as the maximum potential crop yield in calories per hectare per year. It is constructed based only on geographical characteristics and not actual historical yields which makes suitable as an exogenous control variable. We compute the average of caloric suitability across smaller grid cells within each 1 by 1 degree grid cell. Since the types of crops available in different regions across the globe changed markedly after the Columbian Exchange, Galor and Özak (2016) provide two types of variables, one for the pre-1500 era and one for the post-1500 era. We use the first measure for regressions pertaining to the period prior to 1500 AD and the second measure for the regressions pertaining to the modern period.

Agricultural suitability is an index reflecting the suitability of the climate and geography for agriculture computed by Ramankutty et. al. (2002). The index takes on values between 0 and 1 indicating the probability of cultivation. It is computed as follows: First, the relationship between actual cultivation and exogenous data on soil and climate is estimated in a statistical model. Then the probability of cultivation is predicted for each grid cell based on climatic and geographical characteristics. The index is available across 0.5 by 0.5 decimal degree grid cells. We average this to the level of 1 by 1 degree grid cells.

Distance to nearest river is computed using the location of rivers provided by CIA World Databank II. We compute the distance in km. from the centroid of each 1 by 1 degree grid cell to the nearest major river.

Distance to coast is based on the location of shorelines from Natural Earth (2017). We compute the distance in km. from the centroid of each 1 by 1 degree grid cell to the nearest coast.

Time since Neolithic transition is the number of years elapsed since the earliest evidence of agriculture in the grid cell. We compute it based the list of archaeological Neolithic sites compiled by Pinhasi et al (2005) who use it to trace the spread of the transition from hunting and gathering to agriculture throughout Europe and the Middle East. Each site is

[2] A buffer of 5 km for an oppidum is broadly consistent an area of about 7000 ha of agricultural land, outside a wall of 6 km lenght securing an area of 300 ha.

provided with a radio-carbon date measured in years before present. We define the time since the Neolithic transition as the years since the earliest Neolithic site within each 1 by 1 degree grid cell.

Ancient mines. We use the Digital Atlas of the Roman Empire (DARE) database constructed by Åhlfeldt (2017) to compute the number of ancient mines in each cell. This variable is used as a control to ensure the estimated effect of roads on economic activity is not confounded by mineral deposits that were constructed before or simultaneously with the roads. For each country-pixel, we count the number of mines that existed prior to year 500 AD.

Caravan routes is based on the georeferenced map of the historical caravan network from Lado et. al. (2020). Using the same methodology as in the case of the Roman and modern road network, we compute the fraction of 5 km. buffer zones around the caravan network that lies within each country-cell.

Distance to coal is computed as the distance to geological strata from the carboniferous era from Asch (2005). Fernihough and Rourke (2021) use proximity to the carboniferous era as an instrument for proximity to historical coal fields and find a positive effect on the growth of European cities during the industrial revolution.

References

[1] Andersen, T. B., Dalgaard, C. J., & Selaya, P. (2016). Climate and the emergence of global income differences. *Review of Economic Studies*, 83(4), 1334-1363.

[2] Asch, K. (2005). Internationale Geologische Karte von Europa und den angrenzenden Regionen 1:5.000.000 (IGME5000). Downloaded 11-15-2021 from https://produktcenter.bgr.de/

[3] Fernihough, A. and O'Rourke, K. H. (2021). Coal and the European industrial revolution. *The Economic Journal* 131(635): 1135-1149

[4] Galor, O. and Ö. Özak. (2016) The agricultural origins of time preference. *The American Economic Review* 106(10): 3064-3103.

[5] Lado, S., Elbers,J. P., Doskocil, A., Scaglione, D., Trucchi, E. Banabazi, M. H., Almathen, F., Saitou, N. Ciani, E. and Burger, P.A. (2020). "Genome-wide diversity and global migration patterns in dromedaries follow ancient caravan routes." *Communications biology* 3(1): 1-8.

[6] Natural Earth (2017) http://www.naturalearthdata.com/

[7] New, M., D. Lister, M. Hulme and I. Makin (2002). A high-resolution data set of surface climate over global land areas. *Climate research*, 21(1), 1-25.

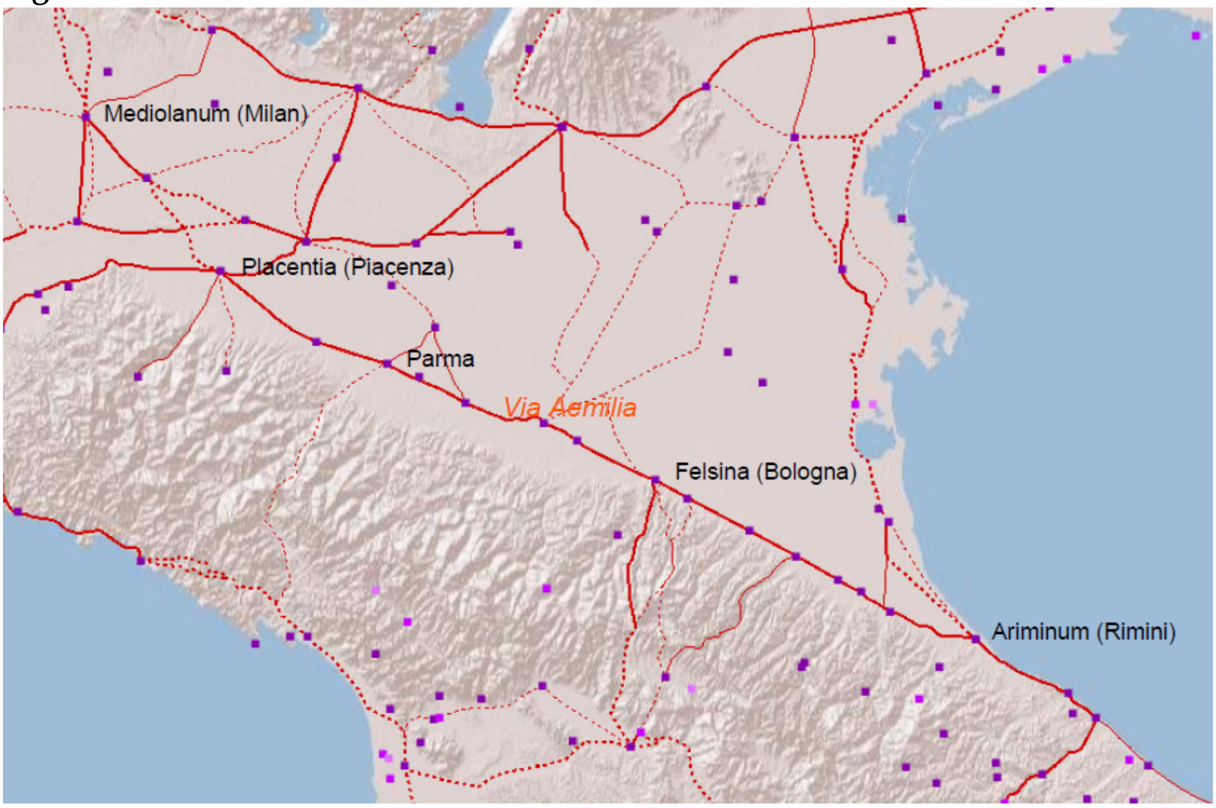
[8] Nunn, N., and D. Puga (2012). Ruggedness: The blessing of bad geography in Africa. Review of Economics and Statistics, 94(1), 20-36.

[9] Pinhasi, R., J. Fort, and A. Ammerman (2005) "Tracing the Origins and Spread of Agriculture in Europe" *PLOS Biology* 3(12), 2220-2228.

[10] Åhlfeldt, J. (2017) *Digital Atlas of the Roman Empire*. Web publication available at http://dare.ht.lu.se/. Lund University, Lund.

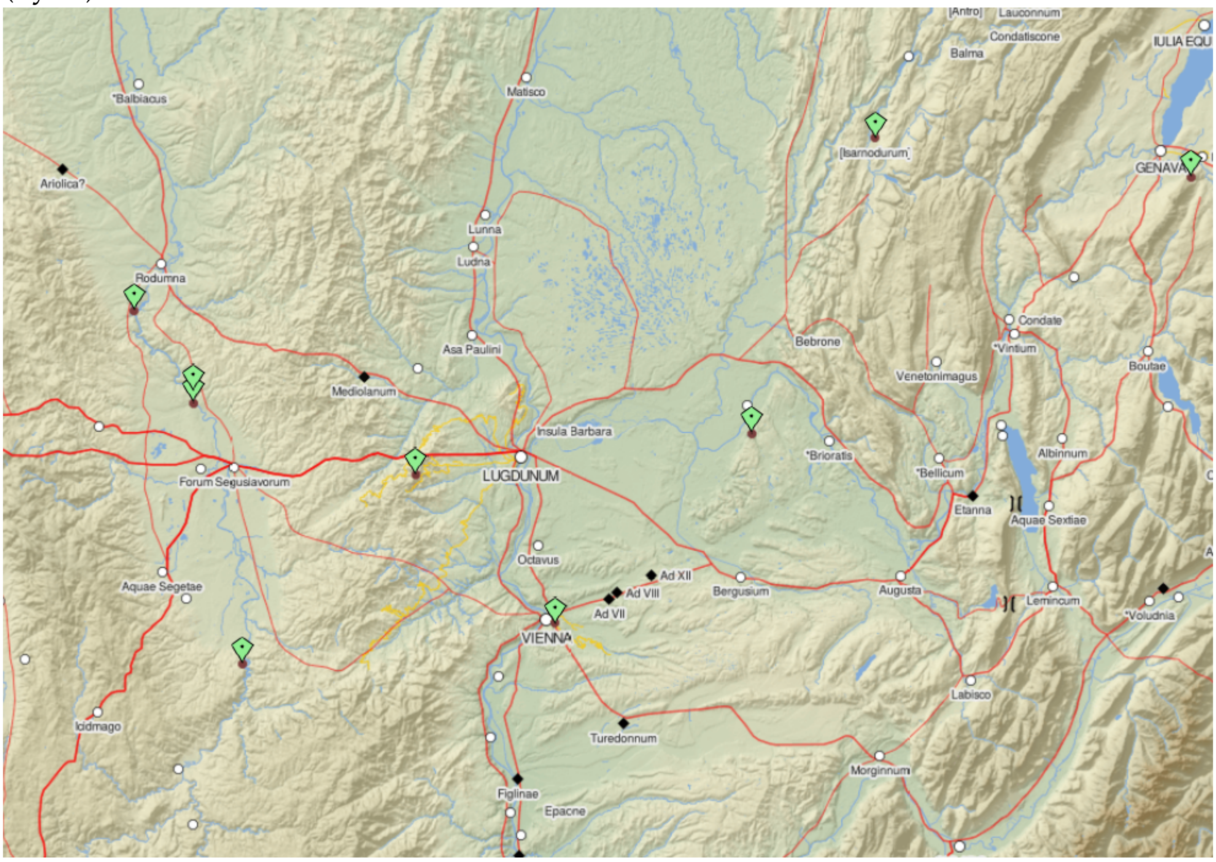
Appendix figures

Figure A1: Via Aemilia from Ariminum to Placentia in 187 BCE



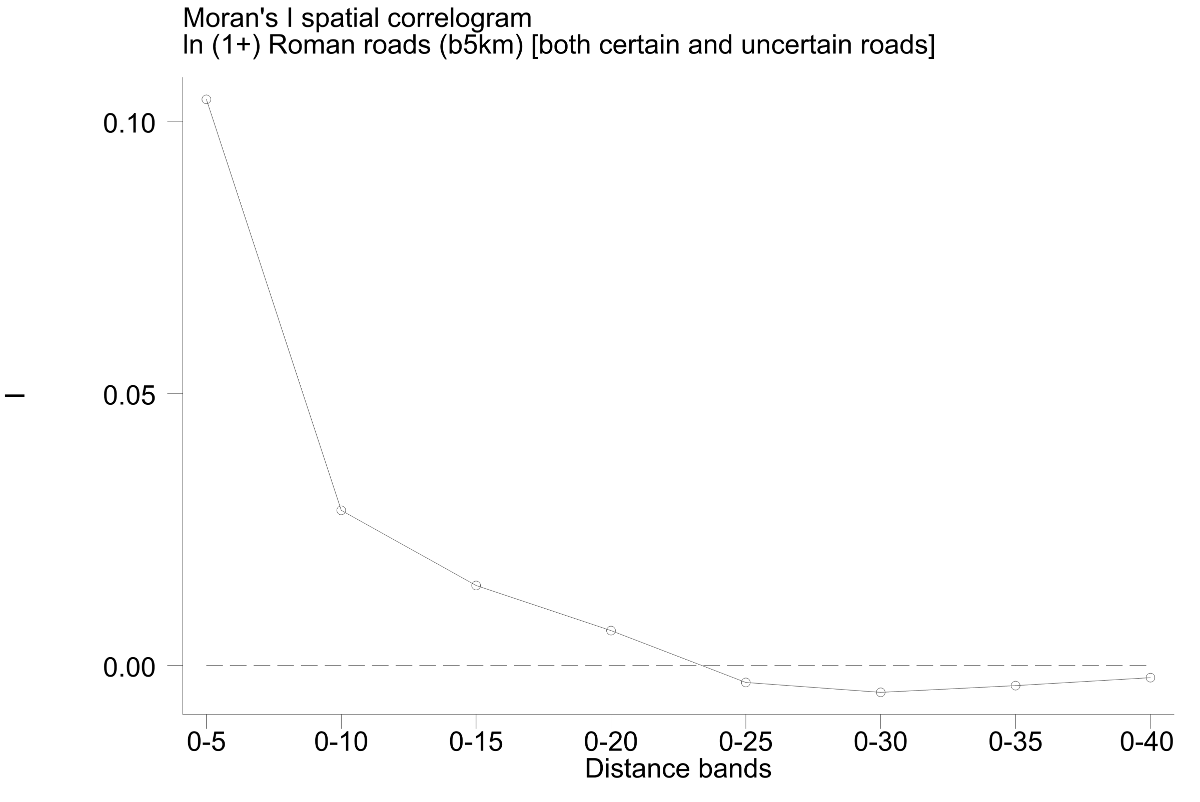
Note: The figure shows Via Aemilia from Ariminum to Placentia, completed in 187 BCE. It also shows other later confirmed roads as thick red lines and unconfirmed roads as dotted lines. Quadratic symbols denote later Roman towns and settlements during antiquity. Source: Created on the basis of data in Talbert (2000).

Figure A2: Roman towns, pre-Roman towns and roads emanating from Lugdunum (Lyon) around 20 BCE



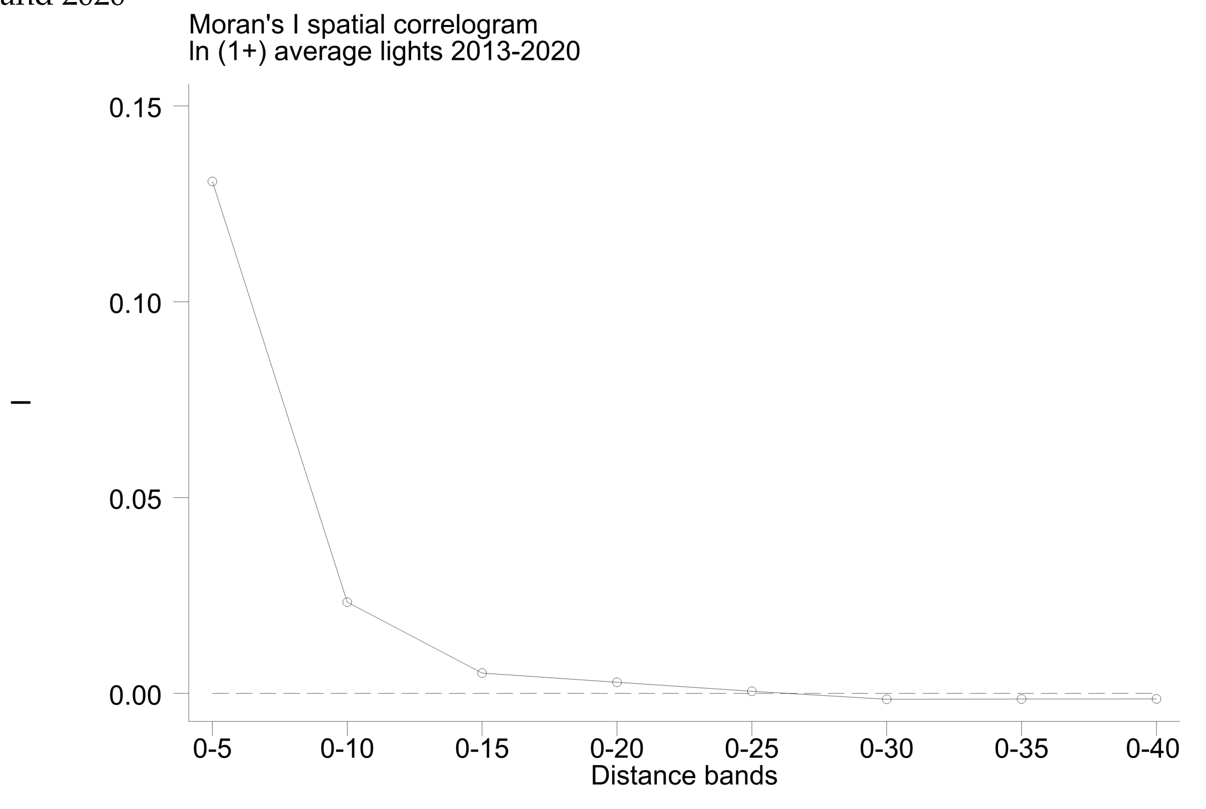
Note: The figure shows Roman towns and settlements after Caesar's conquest in 58-50 BCE as dark and white circles and pre-Roman Celtic settlements (oppida) as green arrows. Red lines show Roman roads, including Via Agrippa from hub city Lugdunum (Lyon) south along the Rhône. Source: Created on the basis of data in Åhlfeldt (2015).

Figure A3: Spatial correlogram for the density of Roman roads [buffer of 5 km around certain and uncertain roads



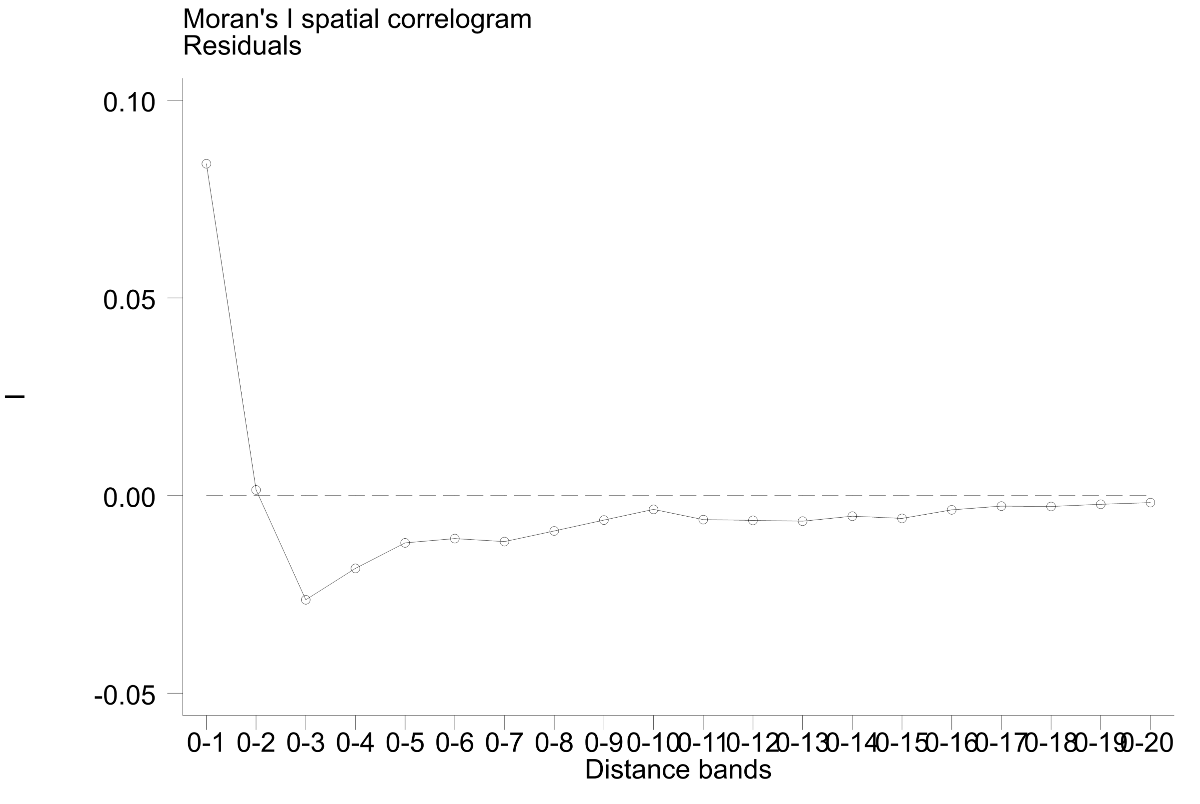
Note: The spatial correlogram displays Moran's I for different distance bands, from cumulative bands of 0 to 40 neighboring grid cells of 1 degree latitude by 1 degree longitude, in increments of 5.

Figure A4: Spatial correlogram for the intensity of average nightlights between 2013 and 2020



Note: The spatial correlogram displays Moran's I for different distance bands, from cumulative bands of 0 to 40 neighboring grid cells of 1 degree latitude by 1 degree longitude, in increments of 5.

Figure A5: Spatial correlogram for the residual of a regression of the intensity of night-lights (average 2013-2020) on the density of Roman roads, controlling for geographic controls, country and language area fixed effects



Note: The spatial correlogram displays Moran's I for different distance bands, from cumulative bands of 0 to 10 neighboring grid cells of 1 degree latitude by 1 degree longitude, in increments of 1.

Appendix Tables

Table A1: Descriptive Statistics

	Mean	SD	Min	Max	N
ln (1+) Roman roads	0.235	0.157	0.000	0.693	693
ln (1+) Modern roads - ESRI (new data)	0.298	0.163	0.000	0.693	693
Roman Settlements in 500 CE	0.574	0.672	0.000	3.367	693
ln (1+) Nightlights from 2013-2020	0.806	0.655	0.000	4.630	691
ln Population in 2010	12.293	2.104	0.000	16.168	693
ln (50+) Elevation	5.931	0.882	3.892	7.933	693
ln Post-1500 caloric suitability	8.431	2.089	0.000	9.621	693
ln Distance to coast	4.087	1.489	-2.580	6.231	693
ln Distance to major river	4.238	1.586	-2.208	7.101	693
ln Distance to natural harbor	4.890	1.501	0.000	6.711	693
ln Distance to Roman Empire border	6.030	1.509	-2.083	7.717	693
ln (1+) Number of mines	0.273	0.567	0.000	2.890	693
ln Distance to capital	5.333	0.962	-3.094	7.024	693
ln (1+) Distance to coal	4.192	1.508	0.000	6.859	693
ln Distance to university	5.646	1.277	1.402	7.724	693
ln Distance to Wittenberg	7.190	0.543	5.769	8.165	693

Table A2: Correlation matrix of control variables

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
ln (50+ Elevation)	1										
ln Post-1500 caloric suitability [2]	-0.0943*	1									
ln Distance to coast [3]	0.344***	-0.00138	1								
ln Distance to major river [4])	0.0706	-0.137**	-0.447***	1							
ln Distance to natural harbor [5])	0.0131	0.0634	0.470***	-0.318***	1						
ln Distance to Roman Empire border [6])	0.240***	-0.262***	-0.363***	0.475***	-0.133**	1					
ln (1+) Number of mines [7])	0.127**	0.0786	0.137**	-0.0238	0.108*	0.0894*	1				
ln Distance to capital [8])	0.146***	-0.138**	0.00421	0.191***	-0.0530	0.148***	0.119**	1			
ln Distance to coal [9])	-0.0968*	-0.180***	-0.126**	0.171***	0.00916	0.232***	-0.162***	-0.127**	1		
ln Distance to university [10])	0.138**	-0.362***	-0.0847	0.293***	-0.0860*	0.222***	-0.0237	-0.0187	0.337***	1	
ln Distance to Wittenberg [11])	0.0800	-0.341***	-0.322***	0.422***	-0.0991*	0.644***	0.0294	0.0380	0.387***	0.662***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: Determinants of Roman Road intensity: Intensive margin

	Dependent variable: Roman roads						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln Latitude	0.051						-0.169**
	(0.050)						(0.072)
ln (10+) Longitude	0.062**						0.118***
	(0.031)						(0.038)
ln Longitude squared	-0.005						-0.021**
	(0.006)						(0.009)
ln (50+) Elevation	-0.078***						-0.038***
	(0.007)						(0.008)
ln Pre-1500 caloric suitability	0.017***						0.015***
	(0.003)						(0.002)
Africa dummy		-0.041**					-0.008
		(0.019)					(0.026)
Middle East dummy		-0.039***					0.005
		(0.015)					(0.026)
ln Distance to coast			-0.060***				-0.043***
			(0.005)				(0.006)
ln Distance to major river			-0.015***				-0.016***
			(0.004)				(0.005)
ln Distance to natural harbor			0.021***				0.014***
			(0.004)				(0.004)
ln Distance to Roman Empire border			-0.040***				-0.032***
			(0.005)				(0.008)
ln (1+) Number of mines				-0.013			0.008
				(0.008)			(0.008)
ln (1+) Oppida (b5km)					0.557		
					(0.606)		
ln Years since Neolithic transition						0.003	
						(0.059)	
Observations	693	693	693	693	183	183	693
R^2	0.260	0.014	0.294	0.002	0.004	0.000	0.351

Notes: Roman roads are defined as the log of one plus the fraction of a 5 km buffer around the Roman road system that lies wihin the total area of a country-cell. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A4: Robustness $\frac{1}{2}$ by $\frac{1}{2}$ Grid Cells

	Modern roads	Settlements	Lights	Pop. den.
	(1)	(2)	(3)	(4)
ln (1+) Roman roads	0.164***	0.480***	0.759***	1.011***
	(0.032)	(0.073)	(0.090)	(0.192)
ln Area	0.025***	0.104***	0.053***	1.192***
	(0.006)	(0.010)	(0.017)	(0.032)
ln (50+) Elevation	-0.034***	-0.021	-0.201***	-0.351***
	(0.008)	(0.018)	(0.024)	(0.058)
ln Post-1500 caloric suitability	0.012***		0.007	0.187***
	(0.003)		(0.009)	(0.025)
ln Distance to coast	-0.016***	-0.026*	-0.124***	-0.232***
	(0.006)	(0.013)	(0.020)	(0.040)
ln Distance to major river	-0.008*	-0.021*	-0.026**	-0.121***
	(0.004)	(0.011)	(0.013)	(0.033)
ln Distance to natural harbor	-0.002	-0.039***	-0.039**	-0.127***
	(0.005)	(0.013)	(0.019)	(0.035)
ln Distance to Roman Empire border	-0.008	0.011	0.003	0.013
	(0.009)	(0.013)	(0.025)	(0.053)
ln (1+) Number of mines	-0.013	0.098***	0.008	-0.024
	(0.010)	(0.030)	(0.030)	(0.061)
ln Distance to capital	-0.021***		-0.233***	-0.398***
	(0.007)		(0.032)	(0.056)
ln (1+) Distance to coal	-0.004		-0.020**	-0.051**
	(0.003)		(0.010)	(0.022)
ln Distance to university	0.000		-0.043	-0.031
	(0.007)		(0.029)	(0.054)
ln Distance to Wittenberg	-0.014		-0.266**	-0.305
	(0.038)		(0.113)	(0.243)
ln Pre-1500 caloric suitability		0.033***		
		(0.007)		
Country FE	Yes	Yes	Yes	Yes
Country-language FE	Yes	Yes	Yes	Yes
Observations	1785	1785	1781	1785
R^2	0.416	0.263	0.569	0.735

Notes: Roman roads and modern roads are defined as the log of one plus the fraction of a 5 km buffer around, respectively, the Roman and modern road system that lies wihin the total area of a country-cell. Roman settlements is log of one plus the number of major settlements within the country-cell in 500 CE. Nightlights is log of the average light intensity measured at night by satelite from 2013-2020 CE. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A5: Sample Split and $\frac{1}{2}$ by $\frac{1}{2}$ Grid Cells

Dependent variable:	Settlements in 500 CE		Modern roads		Nightlights from 2013-2020		Population in 2010	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	Europe	MENA	Europe	MENA	Europe	MENA	Europe	MENA
ln (1+) Roman roads	0.413^{***}	0.636^{***}	0.192^{***}	0.081	0.745^{***}	0.741^{***}	1.148^{***}	0.563
	(0.081)	(0.154)	(0.038)	(0.064)	(0.105)	(0.168)	(0.207)	(0.426)
ln Area	0.086^{***}	0.183^{***}	0.026^{***}	0.023^{**}	0.057^{***}	0.063^{*}	1.184^{***}	1.320^{***}
	(0.010)	(0.026)	(0.007)	(0.011)	(0.020)	(0.033)	(0.035)	(0.074)
ln (50+) Elevation	-0.031	0.002	-0.049^{***}	-0.003	-0.245^{***}	-0.169^{***}	-0.467^{***}	-0.233
	(0.022)	(0.040)	(0.009)	(0.015)	(0.029)	(0.048)	(0.053)	(0.143)
ln Pre-1500 caloric suitability	0.016^{*}	0.036^{***}						
	(0.009)	(0.010)						
ln Distance to coast	-0.003	-0.091^{***}	-0.011	-0.033^{***}	-0.117^{***}	-0.160^{***}	-0.178^{***}	-0.440^{***}
	(0.014)	(0.031)	(0.008)	(0.010)	(0.026)	(0.036)	(0.041)	(0.098)
ln Distance to major river	-0.008	-0.048^{*}	-0.002	-0.020^{**}	0.005	-0.120^{***}	-0.017	-0.450^{***}
	(0.012)	(0.027)	(0.004)	(0.009)	(0.014)	(0.032)	(0.027)	(0.118)
ln Distance to natural harbor	-0.015	-0.058^{**}	-0.006	-0.004	-0.018	-0.131^{***}	-0.050	-0.324^{***}
	(0.017)	(0.023)	(0.007)	(0.009)	(0.024)	(0.033)	(0.041)	(0.081)
ln Distance to Roman Empire border	0.027^{**}	0.033	-0.006	-0.048	0.004	-0.207	0.031	-0.006
	(0.013)	(0.113)	(0.009)	(0.081)	(0.026)	(0.294)	(0.054)	(0.792)
ln (1+) Number of mines	0.099^{***}	0.054	-0.014	-0.002	0.008	-0.024	-0.015	-0.195
	(0.032)	(0.094)	(0.011)	(0.023)	(0.031)	(0.094)	(0.061)	(0.221)
ln Post-1500 caloric suitability			0.012^{**}	0.011^{***}	0.009	0.005	0.132^{***}	0.192^{***}
			(0.005)	(0.004)	(0.014)	(0.013)	(0.032)	(0.037)
ln Distance to capital			-0.017^{*}	-0.032^{***}	-0.207^{***}	-0.365^{***}	-0.321^{***}	-0.716^{***}
			(0.009)	(0.011)	(0.044)	(0.049)	(0.072)	(0.104)
ln (1+) Distance to coal			-0.006	-0.002	-0.022^{**}	-0.007	-0.043^{**}	-0.036
			(0.004)	(0.007)	(0.011)	(0.028)	(0.021)	(0.082)
ln Distance to university			0.000	-0.072	-0.058^{**}	0.802^{***}	-0.067	1.354^{***}
			(0.008)	(0.062)	(0.029)	(0.216)	(0.053)	(0.463)
ln Distance to Wittenberg			-0.051	0.290	-0.457^{***}	0.474	-0.645^{***}	0.681
			(0.042)	(0.251)	(0.125)	(0.838)	(0.232)	(2.494)
Country-language FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1216	569	1216	569	1212	569	1216	569
R^2	0.267	0.246	0.388	0.332	0.583	0.588	0.811	0.665

Notes: Roman roads and modern roads are defined as the log of one plus the fraction of a 5 km buffer around, respectively, the Roman and modern road system that lies wihin the total area of a country-cell. Roman settlements is log of one plus the number of major settlements within the country-cell in 500 CE. Nightlights is log of the average light intensity measured at night by satelite from 2013-2020 CE. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A6: Only certain roads

	Modern roads	Settlements	Lights	Pop. den.
	(1)	(2)	(3)	(4)
ln (1+) Roman roads (only certain)	0.265***	0.101	0.707***	0.989***
	(0.071)	(0.252)	(0.198)	(0.373)
ln Area	0.026***	0.194***	0.071**	1.145***
	(0.010)	(0.026)	(0.030)	(0.036)
ln (50+) Elevation	-0.060***	-0.132***	-0.253***	-0.509***
	(0.014)	(0.050)	(0.042)	(0.080)
ln Post-1500 caloric suitability	0.008		-0.000	0.168***
	(0.007)		(0.022)	(0.034)
ln Distance to coast	-0.009	-0.014	-0.080**	-0.099*
	(0.011)	(0.030)	(0.036)	(0.052)
ln Distance to major river	0.007	-0.032	0.008	0.009
	(0.007)	(0.032)	(0.024)	(0.055)
ln Distance to natural harbor	-0.000	-0.084***	-0.027	-0.035
	(0.007)	(0.027)	(0.023)	(0.042)
ln Distance to Roman Empire border	0.001	-0.009	-0.011	0.079
	(0.013)	(0.034)	(0.043)	(0.076)
ln (1+) Number of mines	-0.005	0.159***	0.008	-0.024
	(0.009)	(0.052)	(0.031)	(0.061)
ln Distance to capital	-0.000		-0.188***	-0.339***
	(0.011)		(0.055)	(0.086)
ln (1+) Distance to coal	-0.007		-0.037**	-0.094***
	(0.005)		(0.017)	(0.032)
ln Distance to university	0.003		-0.028	0.004
	(0.011)		(0.044)	(0.082)
ln Distance to Wittenberg	-0.101*		-0.467**	-0.925***
	(0.057)		(0.188)	(0.327)
ln Pre-1500 caloric suitability		0.062*		
		(0.032)		
Country FE	Yes	Yes	Yes	Yes
Country-language FE	Yes	Yes	Yes	Yes
Observations	522	522	521	522
R^2	0.540	0.487	0.689	0.864

Notes: Roman roads and modern roads are defined as the log of one plus the fraction of a 5 km buffer around, respectively, the Roman and modern road system that lies wihin the total area of a country-cell. Roman settlements is log of one plus the number of major settlements within the country-cell in 500 CE. Nightlights is log of the average light intensity measured at night by satelite from 2013-2020 CE. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A7. Only certain roads and Sample Split

Dependent variable:	Settlements in 500 CE		Modern roads		Nightlights in 2010		Population in 2010	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	Europe	MENA	Europe	MENA	Europe	MENA	Europe	MENA
ln (1+) Roman roads (only certain)	0.335	-0.057	0.308^{***}	0.142	0.896^{***}	0.086	1.543^{***}	-1.128
	(0.273)	(0.779)	(0.075)	(0.239)	(0.205)	(0.729)	(0.395)	(1.188)
ln Area	0.215^{***}	0.279^{***}	0.032^{***}	-0.048^{**}	0.090^{***}	-0.061	1.174^{***}	1.126^{***}
	(0.027)	(0.091)	(0.010)	(0.024)	(0.033)	(0.078)	(0.039)	(0.121)
ln (50+) Elevation	-0.128^{**}	-0.263^{*}	-0.068^{***}	-0.009	-0.233^{***}	-0.332^{***}	-0.441^{***}	-0.929^{***}
	(0.058)	(0.139)	(0.015)	(0.027)	(0.046)	(0.104)	(0.094)	(0.233)
ln Pre-1500 caloric suitability	-0.035	0.104^{***}						
	(0.026)	(0.039)						
ln Distance to coast	0.003	-0.088	0.007	-0.021	-0.082^{*}	-0.043	-0.122^{**}	-0.016
	(0.030)	(0.098)	(0.013)	(0.020)	(0.046)	(0.078)	(0.055)	(0.157)
ln Distance to major river	-0.014	-0.115	0.014^{**}	-0.026^{*}	0.028	-0.106^{*}	0.052	-0.220
	(0.029)	(0.139)	(0.007)	(0.015)	(0.023)	(0.058)	(0.048)	(0.174)
ln Distance to natural harbor	-0.058^{*}	-0.047	-0.008	-0.003	-0.014	-0.084^{*}	-0.012	-0.017
	(0.032)	(0.051)	(0.009)	(0.012)	(0.032)	(0.044)	(0.054)	(0.087)
ln Distance to Roman Empire border	0.007	-0.518	0.002	-0.511^{**}	-0.013	-1.047	0.074	-0.047
	(0.034)	(0.598)	(0.013)	(0.218)	(0.044)	(0.858)	(0.079)	(1.449)
ln (1+) Number of mines	0.095^{*}	0.445^{**}	-0.014	0.059^{**}	-0.022	0.167	-0.082	0.226
	(0.050)	(0.174)	(0.009)	(0.029)	(0.033)	(0.120)	(0.061)	(0.265)
ln Post-1500 caloric suitability			0.012^{*}	0.006	0.007	0.009	0.130^{***}	0.179^{***}
			(0.007)	(0.009)	(0.034)	(0.029)	(0.048)	(0.064)
ln Distance to capital			-0.012	-0.028	-0.183^{***}	-0.363^{***}	-0.326^{***}	-0.433^{**}
			(0.012)	(0.020)	(0.065)	(0.093)	(0.102)	(0.192)
ln (1+) Distance to coal			-0.009	-0.015	-0.042^{**}	0.004	-0.104^{***}	0.243
			(0.005)	(0.015)	(0.018)	(0.052)	(0.031)	(0.171)
ln Distance to university			0.007	0.036	-0.031	0.576	-0.012	0.026
			(0.011)	(0.108)	(0.044)	(0.422)	(0.080)	(0.894)
ln Distance to Wittenberg			-0.110^{*}	1.169^{**}	-0.535^{***}	2.582	-1.004^{***}	-0.718
			(0.057)	(0.526)	(0.197)	(1.956)	(0.322)	(3.475)
Country-language FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	403	119	403	119	402	119	403	119
R^2	0.459	0.545	0.538	0.639	0.692	0.742	0.885	0.827

Notes: Roman roads and modern roads are defined as the log of one plus the fraction of a 5 km buffer around, respectively, the Roman and modern road system that lies wihin the total area of a country-cell. Roman settlements is log of one plus the number of major settlements within the country-cell in 500 CE. Nightlights is log of the average light intensity measured at night by satelite from 2013-2020 CE. The analysis is performed on country-cells within the Roman empire containing non-zero Roman roads. Heteroskedasticity robust standard errors are reported in parentheses, Conley standard errors in square brackets. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.