The Road to Riches:
Quantifying the Persistent Effects of the Silk Roads Network

A Thesis
Presented to
The Faculty of the Department of Economics
Bates College

In partial fulfillment of the requirements for the
Degree of Bachelor of Arts

by

Zofia Ahmad
Lewiston, Maine
March 20, 2019

Acknowledgements

I would first like to thank my advisor, Luke Chicoine, for his unerring advice, constant encouragement, and everlasting patience. I would also like to thank my parents for their abiding love and support, and my brother for his (un)helpful motivational speeches. Finally, to Sierra, Becca, Kanchan, and Chris: thank you for your friendship and willingness to listen to me ramble on about grid cells and camels for hours with no more than the requisite number of Google Maps jokes. Without you the road to the completion of this thesis would have been far shorter, but not nearly so enjoyable. And speaking of roads…

Table of Contents

Introduction

In 2013, Chinese President Xi Jinping announced a massive new trade and transportation project entitled the 'One Belt, One Road Initiative.' By explicitly reviving the Silk Road, Xi invoked the strong cultural memory of one of the most enduring and influential trade networks in world history. The regions along the Silk Road network were exposed to transcontinental trade for over 2,000 years, forming an intriguing natural experiment of the endurance of trade and transportation networks. Qualitative evidence suggests the route had a plethora of cultural and economic effects on the surrounding areas, however there has not yet been a quantitative study on the persistence and significance of these effects.

In this thesis, I study the persistence of the Silk Road trading network over a period of about 2,000 years, both in terms of path dependence throughout its history, and economic activity in the modern day. Trade over the Silk Road network can be split into three peak periods, each aligning with one or several powerful empires that ensured stability along the routes: ca. 100 B.C. to A.D. 200 (Roman Empire and Han Dynasty), 675-875 (Tang Dynasty and Umayyad Caliphate) and 1245-1335 (Mongol Empire).[1] In this project I use digitized maps of the routes during these peak periods, in conjunction with modern-day transportation networks and satellite images, to analyze the effect of the Silk Road on contemporary economic activity levels, population density and path dependence along the historical routes. Contemporary economic activity is measured using estimates of GDP based on nighttime light intensity.[2]

---

[1] Stephan Barisitz, *Central Asia and the Silk Road: Economic Rise and Decline over Several Millennia*, Studies in Economic History Ser (New York: Springer, 2017), http://LW5CZ6WA6G.search.serialssolutions.com/?V=1.0&L=LW5CZ6WA6G&S=JCs&C=TC 027855929&T=marc&tab=BOOKS.

[2] J. Vernon Henderson, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *The American Economic Review* 102, no. 2 (2012): 994–1028.

I find that past proximity to a Silk Route in any period leads to significantly higher contemporary levels of nighttime light intensity and population density, as well as a higher likelihood of major highway presence and closer proximity to a major highway. I find that the magnitudes of these results are progressively lower for older periods, indicating that the effects fade over time. Furthermore, these results are consistently larger for areas closer to the routes, and progressively fade away as distance to the route grows higher. The relationship of the magnitude of coefficients for all periods remains consistent through this proximity analysis, indicating that the relationship between route presence and the variables of interest fades over both space as well as time. Presence during the medieval period is associated with a 5.19 percent increase in GDP per capita, while presence during the Tang-Caliphate period is associated with a 4.74 percent increase and presence of a route in an ancient period is associated with a 4.02 increase.[3]

Studying the persistence of the Silk Road over such a long time span allows for several advantages. Most studies of persistence understandably use institutions from the late modern period (ca. 18th century to the present) due to the lack of concrete bodies of information about institutions in the ancient past. The wealth of pre-existing cartographic data on the Silk Road system permits a much further-reaching analysis of path dependence and persistence. The GIS techniques used to examine the evolution and persistence of the Silk Road allow me to use a combination of geospatial and quantitative methods to study the importance of an institution that was formed more than two thousand years ago. The length of the study allows for a deeper understanding of how transportation networks persist through extremely long periods of time. Furthermore, the scope of this project allows me to look at evolutionary persistence from period to period up until the present, exploring the contours of change in a way that is rarely possible.

[3] These are calculated following Dickens 2017, who writes that the percentage change in GDP per capita ≈ percentage change in night lights x 0.3, assuming that $\ln(0.01 + \text{nightlights}_g) \approx \ln(\text{nightlights}_g)$.

Literature Review

This paper relates to two interconnected bodies of literature. First, my analysis of the degree of cartographic continuity between peak periods of the Silk Road network expands on the growing body of evidence that examines the persistence of spatial equilibria despite the obsolescence of the factors that led to the establishment of those institutions/cities in the first place (Bleakley & Lin, 2012), or in the face of major shocks (Jedwab, Kerby & Moradi, 2017, Michaels & Rauch, 2016, Redding, Sturm & Wolf, 2011). While there is evidence for institutional persistence across countries (Acemoglu, Johnson & Robinson, 2001), the Silk Road network spanned a significantly longer time period than many of the previous studies. In addition, much of the research on path dependence has focused on the persistence of physical places and infrastructures (cities, airports, railroad networks etc.) rather than on the persistence of movement patterns through these locations.

Second, my research on the effect of the Silk Road network on the contemporary GDP of the surrounding regions provides a contribution to the ongoing debate on the effect of long-term trade and infrastructural investment on economic well-being. In this my research has much in common with Berger and Enflo (2014), Jedwab and Moradi (2015), and Cosar and Demir (2016). Berger and Enflo (2014) look at the effect of railroad access on the size of towns in Sweden over a 150-year period, and find that towns close to railroads grew larger, while towns further from railroads experienced a negative spillover effect. Jedwab and Moradi (2015), looking at railroad networks in Kenya, find that the effects of colonial railroad investments on the distribution and aggregate level of economic activity persist to this date, even though the railroad network has collapsed and competing regional road networks have expanded considerably. Cosar and Demir (2016) look at large public investment in roads in Turkey and find that proximity to high-capacity roadways significantly increases local access to international trade. While the actual types of infrastructure discussed in these articles are not directly analogous to the Silk Road trading routes, the principles behind the analysis hold true in terms of increased relative access to higher quality transportation and international trade. The vast majority of quantitative studies on transportation and trade examined institutions that were active within the last two centuries.

There is one notable exception to this. Dalgaard, Kaarsen, Olsson and Selaya (2018) look at the persistence of the provision of public goods through the Roman road network. They find that areas that had a high density of Roman roads have the following features today: higher road density, greater settlement formation in A.D. 500, and greater economic activity in 2010. They argue for strong links between economic activity and continued access to transportation. The Silk Road presents a fascinating complement to the Roman Road network. While both are ancient institutions through which trade and culture flowed, there are several notable differences between the two. The Roman road network was built by a single political entity, motivated entirely by political/militaristic goals for expansion. The Silk Road, on the other hand, was built in a much more piecemeal fashion with trade being the primary motivator. This independence from any one political state means that analysis of the route is a purer analysis on the effect of trade and movement through a variety of different political systems and regions. Additionally, the Roman Road network was a far more physical manifestation of transportation networks, whereas the Silk Road was much less tangibly defined, existed for a longer period of time and was more mutable. Therefore, the study of the Silk Road is better described as an analysis of the effect of persistent patterns of movement and trade, rather than physical infrastructure investments by a political actor.

Background

The Silk Road was not a single road but rather a network of shifting, unmarked trade routes stretching from Eastern China to the Eastern Mediterranean.[4] Indeed, the name 'Silk Road' was not coined until Baron Ferdinand von Richthofen coined the term in 1877.[5] None of the routes were paved or systematically enforced, and use of individual routes varied depending on ecological and political factors. The Silk Road can more accurately be described as a patchwork of routes that connected major oasis towns. These towns functioned as semi-independent city states, the rulers of which strictly oversaw trade and played a major role as buyers of goods/services.[6] The stretches between these city states were essentially wild, so as traders entered each town they would have been passing from unregulated regions to communities in which trade was highly supervised.[7] This was especially true during the Han (206 B.C. — A.D. 220) and Tang dynasties (A.D. 618-907) as both dynasties stationed troops along the routes in Central Asia.[8]

For the purposes of the paper, I have chosen to define the start date of the Silk Road as ca. 200 B.C. This phrasing is slightly misleading as the trade routes discussed have existed in some fashion for significantly longer than 2000 years. I have chosen ca. 200 as the start date due to political changes and a growing desire for silk that occurred during this time period. These developments led to an increase in sustained travel across Central Asia, both east-west and north-south.[9] A large part of this development was due to the Han Dynasty's enforcement of the routes. This 'enforcement' included the posting of garrisons in Central Asia and the requirement of a

---

[4] Valerie Hansen, *The Silk Road: A New History*, p. 5.
[5] Susan Whitfield, *Life along the Silk Road*, p. 1.
[6] Hansen, *The Silk Road*, p. 8.
[7] *ibid.*
[8] *ibid.*
[9] Whitfield, *Life along the Silk Road*, p. 2.

travel pass, or *guosuo*, in order to pass through checkpoints along the route.[10] This type of centralized influence also existed during the 7th through 9th centuries with the combined influence of the Tang Dynasty and the Ummayyid Caliphate, as well as during the 13th and 14th centuries with the Mongol Empire (known in China as the Yuan Dynasty). The routes were variable over time, with changes due to a variety of factors. Some of these changes were highly political — for example, the Han dynasty initiated regular trade around the Taklamakan desert in order to bypass the Xiongnu, with whom they shared persistent enmity.[11]

The Silk Road is often thought of as a network of routes that connected the east (i.e. China) to the west (i.e. Rome), however this description does not reflect the complexity and multiplicity of the network. It certainly connected East Asian and Mediterranean cultures, however Central Asian, Near Eastern, Arabian and even African cultures and peoples also played a major part in the history of the network.[12] Trade was not confined to the East-West dimension. For example, some important sections reached down from Central Asia to India, and upwards to Mongolia and towards Scandinavia. Central Asia in particular was the gravitational center of the network, both geographically and historically.[13] These land routes were complemented by the equally important maritime network, which connected regions as disparate as the Mediterranean to sub-Saharan Africa, India, Southeast Asia and Japan. In addition to the diverse scope of the network, the common conception of caravans traveling directly from Beijing to Rome is misleading for a couple of important reasons.

---

[10] Hansen, *The Silk Road*, p. 17.
[11] *Ibid*, p. 18.
[12] Whitfield, *Life along the Silk Road*, p. 1.
[13] *Ibid*, p. 4.

First, most of the trade that was transported along these routes was low-bulk,[14] and trade generally took the shape of piecemeal 'trickle trade,' with goods passing through many hands before reaching their final destinations.[15] Large long-distance caravans are rarely mentioned, and usually only in regards to state-to-state emissarial exchanges.[16] Most transactions would have involved exchanges of goods rather than of coins.[17] Silk was indeed traded along the route (and sometimes used as currency), however other types of goods, technologies, artistic motifs, ideas and peoples were traded as well. The modes of land travel stayed fairly consistent throughout the time period discussed in this study.[18] Relays of horses were the fastest mode of transport, while camels, yaks, carts, donkeys and foot travel were preferred for different terrains and different levels of wealth.[19]

Second, while there is a great deal of evidence for contact between the Central Asian region and China during the Han Dynasty, the first 'perceptible contacts' between China and the West do not come until the second and third centuries A.D.[20] Prior to this period there is a smattering of vague references to China in Greek and Roman sources (such as in the *Periplus Maris Erythraei*, which details the maritime trade routes of the Indian Ocean); however, there is no material evidence for contact between the two regions.[21] As such, the entirety of the trade during the ancient

---

[14] Barisitz, *Central Asia and the Silk Road: Economic Rise and Decline over Several Millennia*, p. 10.
[15] *Ibid*, p. 37.
[16] Hansen, *The Silk Road*, p. 10.
[17] *Ibid*, p. 4.
[18] Whitfield, *Life along the Silk Road*, p. 200.
[19] *Ibid.*
[20] Hansen, *The Silk Road*, p. 21.
[21] *Ibid*, p. 19. Pliny the Elder, writing in the late first century A.D., does write about silk cloth and silkworms connected with the people of Seres (a land that was unknown to the Romans but was located somewhere along the northern edge of the known world), but this cannot be firmly connected to China in any way.[21] During the ancient time period, silk was also being made in India and the Greek island of Cos (in the eastern Aegean) during this time and it is possible that he was referring to silks from one of these regions. Chinese silks dating back to the first through third centuries have been found in Palmyra, Syria, and are some of the earliest silks to have reached West Asia from China. Furthermore, while there have been thousands of Roman coins found in India during the ancient time period, the earliest European coins found in China are from Byzantium and date to only the 6th century A.D.

(first) period defined in this paper almost certainly included *no* continuous trade between China and Rome.

In addition to the types of trade described above, the Silk Roads also played an important role in the transportation of peoples, belief systems and ideas. While communities along the routes were largely agricultural and did not engage in the trade, those living along the Silk Road played crucial role in transmitting, translating and modifying belief systems as they passed from one civilization to another.[22] Prior to the introduction of Islam, these communities were usually highly tolerant of different faiths and cultures.[23] There is evidence for a high degree of ethnic diversity in the communities living along the Silk Routes.[24] Some of this was certainly due to the migrations of peoples, however communities only absorbed large numbers of refugees due to wars and political unrest.[25] Some of these refugees were impressed into slavery, but many were absorbed into the local population as scribes etc. and contributed to the diversity of the populations living along the Silk Road.[26]

Given the history of the routes, it is thus reasonable to believe that route presence might have had a significant effect on contemporary levels of economic activity, population density and road placement. While one would assume that these effects would be concentrated in the highly-centralized city-states along the routes, I ultimately find evidence that even proximity to a route between sites continues to have a significant effect on contemporary societies. It is important to note that I am focusing on land routes in order to study the effect of the entire routes rather than

[22] Hansen, *The Silk Road*, p. 19.
[23] *Ibid.*
[24] *Ibid*, p. 13.
[25] *Ibid*, p. 4.
[26] *Ibid*, p. 45.

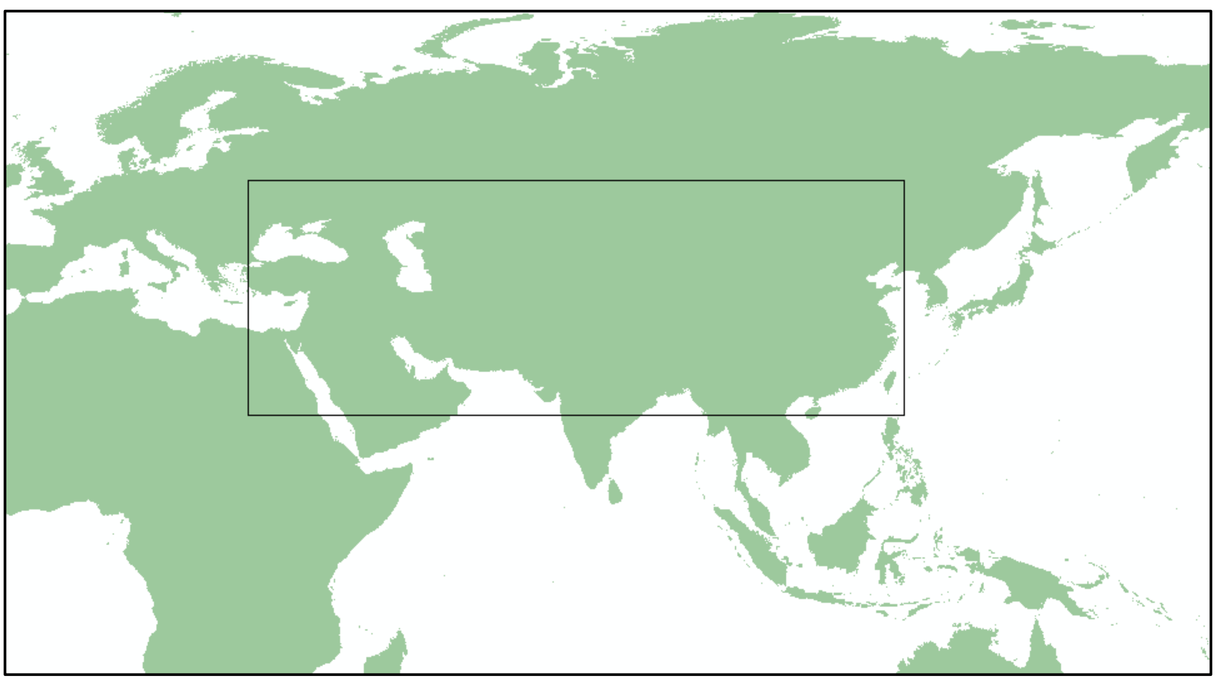
just the cities along the routes, however maritime routes played a major role in the trading network throughout the history of the Silk Road. [27]

Data

Units of Observation

The area under study is defined by the following meridians/parallels: 53 N, 123 E, 19 N and 27 E (as shown in Figure 1). In order to analyze the effect of the Silk Route network on a localized level, the region is divided into 0.5 by 0.5 degree grid cells (as shown in Figure 2) that act as the units of observation. All coordinate systems represent the earth's surface two-dimensionally, however there are several different methods by which to do this. The data used for this paper is assigned to the World Geodetic System (WGS) 1984, the most commonly used of the different systems. The WGS 1984 was created in 1984 and last revised in 2004.[28]

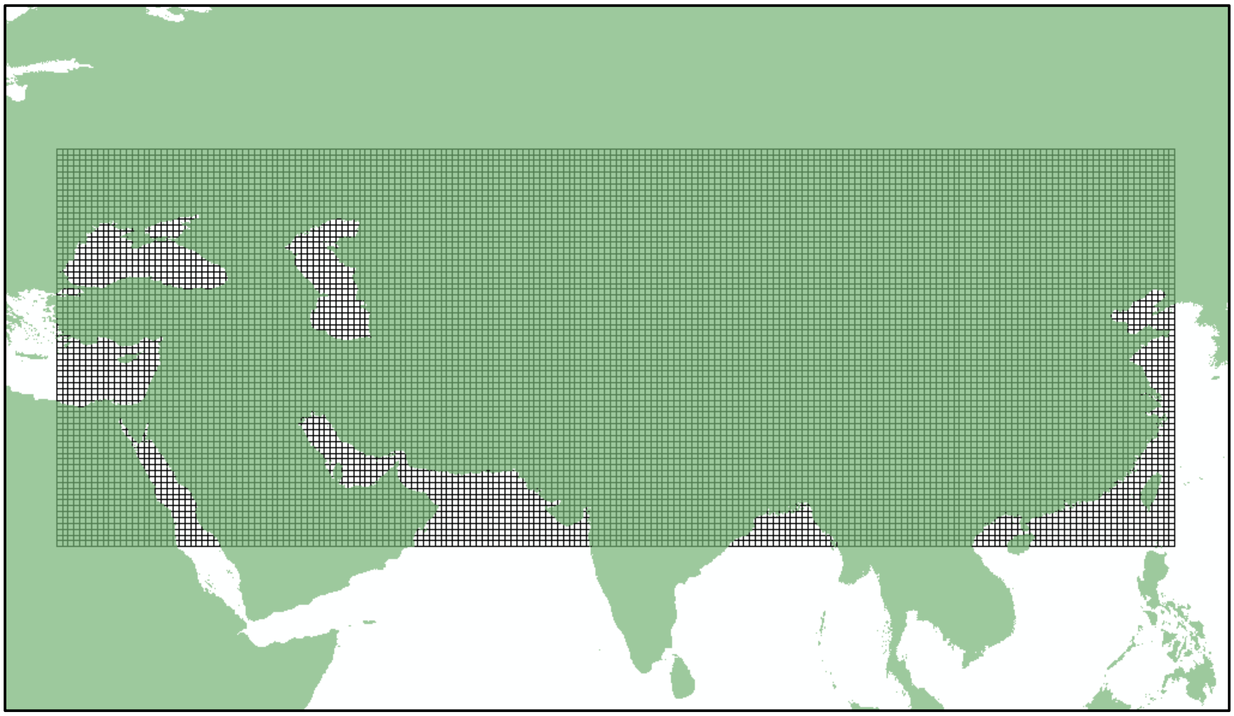
Figure 1: Region of Study



[27] Hansen, *The Silk Road*, p. 45.

[28] Datum: World Geodetic System 1984. *Georepository.* Retrieved from: https://georepository.com/datum_6326/World-Geodetic-System-1984.html

Figure 2: Grid



As described above, grid cell size is 0.5 by 0.5 degrees.[29] The grid cell size used in other papers varies between 0.1 by 0.1 degrees (Jedwab & Moradi 2015, Jedwab & Storeygard 2017) to 1 by 1 degrees (Dalgaard et al. 2018). Because the dimensions of each grid cell are defined by degrees rather than meters, the cell area varies slightly from cell to cell, going to zero as the latitude moves from the equator towards the poles. There were originally 13,056 grid cells in the dataset. Of these, the 1,600 grid cells that included only water (either ocean or a major lake) have been dropped.

[29] I also run robustness checks using 1x1 degree cells. The associated results are shown in Section 5 of the Appendix.

Silk Routes

At the time I was starting the work for this thesis, there did not exist a digitized, spatially-accurate version of the Silk Routes that was also time-specific. In order to analyze the effect of Silk Route presence over time, I cross-referenced several sources to create the first time-specific digital maps of the Silk Routes.[30] Instead, the base map for the ancient and medieval periods is based on a map of the entire network from The Center for Geographic Analysis at Harvard. [31] The file was created by Tim Williams, an Associate Professor in Silk Roads Archaeology at the University College of London. The map is based on archaeological data, and includes routes from throughout the existence of the Silk Roads.[32] In order to convert the base map into time-specific maps, I used the maps of the Silk Roads by time period from Stephen Barisitz's book '*Central Asia and the Silk Road: Economic Rise and Decline over Several Millennia.*' I chose to cross-reference Williams' map with these maps to get a consistent definition of the scope of the Silk Roads for each time period. Barisitz's maps are designed to show directionality rather than specific routes, and thus are impossible to georectify in a reliably accurate fashion, however the map is historically accurate in which cities it connects for each time period.

To create the cross-referenced version of Williams' map, I overlaid the locations of the cities that were connected by the ancient Silk Route network according to Barisitz with the Williams map. I then erased the routes included in Williams' general map that did not connect the

[30] In addition, I also created digital georeferenced maps of the first and third periods based on highly accurate physical maps from two separate sources. I cross-referenced them to Barisitz's definition of the Silk Routes following the same process as described for the Williams map. While I use the cross-referenced and time-specific versions of Williams' maps for the main analysis in the paper, these other routes are used for robustness checks. The process of creating these maps and the results associated with them are described in the first section of the appendix. I was not able to find a map of the routes for the third period that both covered the broad geographical scope necessary for comparative purposes and was sufficiently accurate to allow for georectification.

[31] The Center for Geographic Analysis at Harvard. *Silk Road* [Data file]. Retrieved from: https://worldmap.harvard.edu/maps/7547.

[32] Williams describes the process of creating the map further in his book *The Silk Roads: an ICOMOS Thematic Study.*

sites outlined in the Barisitz map. The routes left were then saved as a separate, ancient routes, shapefile. This process was repeated using the set of cities connected by Barisitiz's maps of the second and third time periods in order to create the corresponding map of the routes for each period. The final products are maps that are as archaeologically accurate as possible, and are edited down to represent Barisitz's definition of the scope of the network in each time period. Out of the 11,456 grid cells used in the analysis for this paper, 509 have at least one route in the first time period, 663 have at least one route in the second time period, and 731 have at least one route in the third time period.

Dependent Variables

Data for nighttime light intensity are collected by the National Oceanic and Atmospheric Administration (NOAA). This is a yearly composite of nighttime light intensity constructed using U.S. Department of Defense weather satellites. The satellites circle the earth 14 times per day and observe every location on the planet every night between 8:30 and 10 p.m. local time. Observations that are contaminated by strong sources of natural light (such as those due to wildfires, auroral activity, the bright half of the lunar cycle, etc.) are removed by NOAA scientists before the data is released publically. The dataset used here is an average of all of these daily values of light intensity for the year 2010. The intensity values are integers that range between 0 (no light) and 63; a handful of values (0.1 percent) are censored at 63.

The literature is somewhat divided on whether to use simple averages of nighttime light intensity[33] or a log transformation of the intensity value.[34] I have chosen to use the log transformation of the intensity value for the implications that this form of the value has been shown

[33] Bleakley & Lin (2012), Hendersen, Storeygard & Weil (2012), Jedwab, Kerby & Moradi (2014).
[34] Dickens (2017), Michalopoulos and Papaioannou (2013, 2014) and Hodler and Raschky (2014)

to have about a locations GDP.[35] The logged transformation is calculated by adding 0.01 to the light intensity value, and taking the natural log of the resultant sum. I add 0.01 to account for the 3,049 cells that have a value of zero in the un-transformed dataset. This addition allows us to preserve the sample size, helps to correct for the non-normal structure of the data, and allows for a semi-elasticity interpretation of the benchmark empirical model.[36]

Population density is taken from the Socioeconomic Data and Applications Center (SEDAC) at NASA.[37] The data is a raster image of population density in 2010, adjusted to match 2015 country-level totals. A logged transformation of population density is used, following the same process as the nighttime light intensity transformation described above.

Data on major highway location is taken from Natural Earth's Global Roads database.[38]

Controls

The vector of 'human' controls includes fixed effects for the country the centroid of the grid cell is in,[39] and for the ethnic combination associated with the location of the centroid of each grid cell.[40] The GREG dataset I am using for data on ethnicities identifies three ethnic groups for every point, so a unique id is generated for every combination of groups. Combinations of groups that appear twenty times or less in the dataset are grouped together as an 'other' category, as are countries that appear twenty times or less.

---

[35] Henderson, Storeygard & Weil (2012).

[36] Dickens 2017.

[37] Center for International Earth Science Information Network - CIESIN - Columbia University, and Information Technology Outreach Services - ITOS - University of Georgia. (2013). *Gridded Population of the World (GPW) v4. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)* [Data File]. Retrieved from: http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-adjusted-to-2015-unwpp-country-totals-rev10.

[38] Natural Earth. Roads [Data File]. Retrieved from: https://www.naturalearthdata.com/downloads/10m-cultural-vectors/roads/.

[39] IPUMS International. *World Map* [Data File]. Retrieved from: https://international.ipums.org/international/gis.shtml

[40] Weidmann, Nils B., Jan Ketil Rød and Lars-Erik Cederman (2010). "Representing Ethnic Groups in Space: A New Dataset". Journal of Peace Research, in press.

The vector of geographical controls includes the following variables: (1) distance from the centroid of the grid cell to the nearest coastline,[41] (2) distance from the centroid of the grid cell to the nearest river,[42] (3) absolute latitude,[43] (4) elevation,[44] (5) standard deviation of elevation (as a measure of ruggedness), and (6) the potential vegetation of a grid cell.[45] Data for potential vegetation is taken from the Center for Sustainability and the Global Environment at the Nelson Institute at the University of Wisconsin-Madison. Potential vegetation data is generated by a computer model known as IBIS, developed at the University of Wisconsin, and predicts the vegetation at a given location if human forms of land use had never existed. Potential vegetation types are separated into fifteen categories (including categories such as Savanna, Tropical Evergreen Forest, and Desert).[46]

Endogeneity of Geography

One of the major potential weaknesses with this analysis is the possibility that the routes were originally placed in the most geographically expedient locations, and thus any persistence we find might be due to the preferential geography of these locations rather than the persistence of past route locations. To address this problem, I conduct t-tests comparing the geographical characteristics of the grid cells within 50km of the routes to the cells between 50-100 km away from the routes. If geography did not play a role in the choice of route location, we would expect

---

[41] Natural Earth. Coastline [Data file]. Retrieved from: https://www.naturalearthdata.com/downloads/10m-physical-vectors/

[42] Natural Earth. Rivers + Lake Centerlines [Data file]. Retrieved from: https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines/

[A]bsolute latitude has been shown to be correlated with GDP on both growth and per capita levels (Theil and Galvez, 1995; Irwin and Tervio, 2002; Gallup et al. 1999), and is accordingly included as an additional control variable.

[44] United States Geological Survey. *Topography* [Data file]. Retrieved from: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SRTM_RAMP2_TOPO

[45] Fixed effects are used for potential vegetation in the regression analysis.

[46] A similar form of vegetation measurement, contemporary type of land cover (i.e. urban, barren, woody savanna, snow and ice, etc.), is used as a robustness check in Section 3 of the Appendix. I chose to use potential vegetation for the main set of results as there is more reason to believe that this measure accurately reflects the climate and geography for the majority of the temporal scope of this project.

to see no significant differences between the two categories of grid cells. I test the following geographical characteristics: (1) elevation, (2) standard deviation of elevation, (3) distance to the closest river, (4) distance to coastline, (5) contemporary land cover is any type of forest, (6) contemporary land cover is shrubland, (7) contemporary land cover is grasslands, (8) contemporary land cover is 'barren,' (9) potential vegetation is any type of forest, (10) potential vegetation is shrubland, (11) potential vegetation is grasslands, (12) potential vegetation is tundra, and (13) potential vegetation is desert. The first four categories are measured in levels (with distances to the closest river and coastline being measured in kilometers), while the last nine are measured using a dummy for each category of landcover type.

Tables 1, 2 and 3 show the results of these t-tests for the Medieval, Tang-Caliphate, and Ancient periods, respectively. Columns 1 and 2 contain the averages for areas 0-50 km away from the routes and areas 50-100 km away from the route, respectively. Column 3 indicates the difference between these two means, column 4 the standard error, and column 5 the number of observations in the t-test. I find that the only category that is consistently significantly different is elevation, however the difference for standard deviation in elevation between the two bins is not significant in any period. Additionally, the t-tests for the ancient period suggest that the grid cells 50-100 km away from the routes 2.17 percentage points are more likely to be desert than the grid cells 0-50 km away from the routes; however, they are 5.89 percentage points less likely to be shrubland. The statistical difference in desert location does not persist to the medieval period. To ensure these differences in desert are not driving any of the results, I drop all cells that have desert as their potential vegetation value from the regression analysis.[47]

[47] Estimates with desert cells included can be found in Section 2 of the Appendix.

Table 1: Medieval Geography T-tests

	(1) 0-50 km	(2) 50-100 km	(3) Diff.	(4) Std. Error	(5) Obs.
Elevation	101.0666	107.3219	-6.2553**	2.5068	1,954
SD Elevation	10.1793	9.6785	0.5008	0.4136	1,964
Dist. to River	93.578	98.848	-5.271	4.184	1,986
Dist. to Coast	1,016.920	1,023.755	-6.835	30.993	1,986
Forest (LC)	0.0182	0.0294	-0.0112*	0.0068	1,986
Shrubland (LC)	0.1090	0.1130	-0.0040	0.0142	1,986
Grasslands (LC)	0.3479	0.3774	-0.0295	0.0217	1,986
Barren (LC	0.3215	0.3141	0.0074	0.0210	1,986
Forest (PV)	0.1308	0.1571	-0.0263*	0.0158	1,986
Shrubland (PV)	0.3642	0.3028	0.0614***	0.0213	1,986
Grasslands (PV)	0.0654	0.0497	0.0157	0.0106	1,986
Tundra/Ice (PV)	0.2207	0.2113	0.0094	0.0186	1,986
Desert (PV)	0.0245	0.0328	-0.0082	0.0075	1,986

Notes: I use two separate measures of landcover/climate: contemporary land cover, and potential vegetation. LC denotes a contemporary land cover category, and PV denotes a potential vegetation category. Column 1 indicates the average for cells with a centroid within 0-50 km of the closest medieval (Mongol) route, and column 2 indicates the average for cells with a centroid within 50-100 km of the closest medieval route. Column 3 indicates the difference between the averages in columns 1 and 2, column 4 indicates the standard error, and column 5 indicates the number of observations in the t-test. * p<0.10, ** p<0.05, *** p<0.01.

Table 2: Tang-Caliphate Geography T-tests

	(1) 0-50 km	(2) 50-100 km	(3) Diff.	(4) Std. Error	(5) Obs.
Elevation	96.1053	107.2536	-11.1483***	2.6879	1,728
SD Elevation	9.8261	9.4154	0.4107	0.4458	1,738
Dist. to River	97.392	105.867	-8.475*	4.597	1,760
Dist. to Coastline	1,098.014	1,124,968	-26.954	31.655	1,760
Forest (LC)	0.0143	0.0269	-0.0126*	0.0067	1,760
Shrubland (LC)	0.1246	0.1370	-0.0124	0.0161	1,760
Grasslands (LC)	0.3085	0.3342	-0.0257	0.0224	1,760
Barren (LC	0.3534	0.3688	-0.0153	0.0230	1,760
Forest (PV)	0.0654	0.0819	-0.0166	0.0125	1,760
Shrubland (PV)	0.4157	0.3572	0.0585**	0.0234	1,760
Grasslands (PV)	0.0766	0.0512	0.0254**	0.0118	1,760
Tundra/Ice (PV)	0.2584	0.2522	0.0062	0.0209	1,760
Desert (PV)	0.0194	0.0384	-0.0190**	0.0079	1,760

Notes: I use two separate measures of landcover/climate: contemporary land cover, and potential vegetation. LC denotes a contemporary land cover category, and PV denotes a potential vegetation category. Column 1 indicates the average for cells with a centroid within 0-50 km of the closest Tang-Caliphate period route, and column 2 indicates the average for cells with a centroid within 50-100 km of the closest T-C route. Column 3 indicates the difference between the averages in columns 1 and 2, column 4 indicates the standard error, and column 5 indicates the number of observations in the t-test. * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Han-Roman Geography T-tests

	(1) 0-50 km	(2) 50-100 km	(3) Diff.	(4) Std. Error	(5) Obs.
Elevation	88.5678	98.0107	-9.4429***	2.9383	1,404
SD Elevation	8.7801	8.6288	0.1513	0.4743	1,412
Dist. to River	90.677	95.158	-4.481	4.408	1,434
Dist. to Coastline	939.906	946.730	-6.824	33.574	1,434
Forest (LC)	0.0156	0.0256	-0.0100	0.0075	1,434
Shrubland (LC)	0.1338	0.1611	-0.0274	0.0187	1,434
Grasslands (LC)	0.2818	0.2861	-0.0043	0.0239	1,434
Barren (LC	0.3623	0.3735	-0.0112	0.0255	1,434
Forest (PV)	0.1143	0.1340	-0.0198	0.0174	1,434
Shrubland (PV)	0.3857	0.3268	0.0589**	0.0254	1,434
Grasslands (PV)	0.0623	0.0527	0.0096	0.0124	1,434
Tundra/Ice (PV)	0.2779	0.2666	0.0114	0.0236	1,434
Desert (PV)	0.0130	0.0346	-0.0217***	0.0079	1,434

Notes: I use two separate measures of landcover/climate: contemporary land cover, and potential vegetation. LC denotes a contemporary land cover category, and PV denotes a potential vegetation category. Column 1 indicates the average for cells with a centroid within 0-50 km of the closest ancient route, and column 2 indicates the average for cells with a centroid within 50-100 km of the closest ancient route. Column 3 indicates the difference between the averages in columns 1 and 2, column 4 indicates the standard error, and column 5 indicates the number of observations in the t-test. * p<0.10, ** p<0.05, *** p<0.01.

Empirical Specification

I run the following regression, using OLS with robust standard errors:[48]

$$\theta_g = a + \beta \cdot R_{tg} + \delta X + \gamma Y + \varepsilon_g \qquad (1)$$

Where θ is the value within each cell, g, of either (1) the logged mean of the population density in the cell, (2) the logged mean of the nighttime light intensity of the cell, (3) whether or not a major highway is present in the cell, or (4) logged distance from the grid cell centroid to the closest major highway. R_{tg} is a dummy variable that denotes the presence of a route, with the subscript indicating the time period, t, (either Han-Roman, Tang-Caliphate, or Mongol) and grid cell, g. In addition to the simple dummy for presence of a route, I also test the effect of closer proximity to routes using 'distance bins.' Each grid cell is assigned a value denoting whether the centroid of

[48] Results using clustered standard errors (with respect to country) are included in Section 4 of the Appendix.

the cell is (1) 0-50 km away from the route, (2) 50-100 km away from the route, (3) 100-150 km away from the route, (4)150-200 km away from the route, or (5) 200+ km away from the route. The furthest group, 200 or more kilometers, is used as the reference group.

X is a vector of political controls (country and ethnicity fixed effects). Y is the matrix of geographical controls (elevation, standard deviation of elevation, distance to closest river, distance to coastline, and potential vegetation fixed effects). As mentioned above, the sample size is 11,456 grid cells. I exclude cells whose potential vegetation is in the 'desert' category in order to account for the statistically significant differences in mean likelihood of a location 0-50 km away from a route being a desert and a location 50-100 km away from a route being a desert, as described previously.

Results

Baseline Results

Table 4 shows the relationship between route presence (in each of the three periods) and nighttime light intensity under 6 different specifications. The results from three separate regressions following equation (1) are shown in column 1; one regression is run for each time period, and no controls are used. Column 2 and 3 add to the regression either the vector of political or geographical controls, respectively, with potential vegetation as the form of landcover that is used. Column 4 shows the results of a regression including both political and geographical controls, with contemporary landcover as the measure of landcover used. Column 5 shows the results of a regression including both political and geographical controls, with potential vegetation as the measure of landcover used. As the results of the t-tests described in the previous description indicate that there may be a significant difference in the likelihood of a cell 50-100 km away from the route being desert compared to the cells 0-50 km away from the route, Column 6 shows the results of a regression identical to that in Column 5 except that all squares with 'desert' as their potential vegetation value are dropped out of the analysis. Column 6 is the preferred specification used for the rest of the results in the main body of this paper.[49] The base model indicates that presence of a medieval route corresponds to a nighttime light intensity 17.3% higher than the intensity found in other locations, while presence of a route during the Tang-Caliphate period corresponds to a nighttime light intensity 15.8% higher and presence of an ancient route corresponds to a nighttime light intensity 13.4% higher. The graduated magnitudes of the coefficients suggests that the persistence of the routes fades out over time.

Table 4: Route Presence and Nighttime Light Intensity

	Dependent Variable: Ln (Nighttime Light Intensity)					
	(1)	(2)	(3)	(4)	(5)	(6)
Medieval (ca. 1225 -1335)	0.153^{***}	0.159^{***}	0.203^{***}	0.135^{***}	0.165^{***}	0.173^{***}
	(0.020)	(0.019)	(0.018)	(0.016)	(0.017)	(0.018)
Tang-Caliphate (ca. 675-875)	0.131^{***}	0.149^{***}	0.185^{***}	0.115^{***}	0.148^{***}	0.158^{***}
	(0.021)	(0.019)	(0.019)	(0.016)	(0.018)	(0.019)
Ancient (ca. 100 B.C. – 200 A.D.)	0.124^{***}	0.144^{***}	0.149^{***}	0.092^{***}	0.131^{***}	0.134^{***}
	(0.024)	(0.022)	(0.021)	(0.018)	(0.020)	(0.025)
Political Controls		x		x	x	x
Geo. Controls			x	x	x	x
Potential Vegetation					x	x
Landcover				x		
Desert Dropped						x
N	11,456	11,456	11,282	11,282	11,282	9,094

Notes: Each column in this table indicates three separate regressions: one using a dummy for medieval route presence, one using a dummy for route presence in the Tang-Caliphate period, and one using a dummy for ancient route presence. Column 1 indicates the results of a simple OLS regression of route presence on nighttime light intensity. Column 2 indicates the results of a regression including only the vector of political controls (with fixed effects for the country and ethnicities located at the grid cell centroid) in addition to route presence, while Column 3 indicates the results of a regression including only the vector of geographical controls (with fixed effects for potential vegetation) in addition to route presence. Column 4 includes the results of the regression of route presence including both political and geographical controls but using contemporary landcover instead of potential vegetation, while Column 5 indicates the results of the regression using both sets of controls with potential vegetation rather than contemporary landcover. Finally, column 6 indicates the results of the same regression in column 5 but with desert cells dropped out. This is the preferred specification used for the rest of the paper. Robust standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[49] Results corresponding to the other permutations of the specification shown in Table 1 are included in the appendix.

Table 5 shows the conditional association between presence of a route in each of the three time periods and: (1) logged average nighttime light intensity in the grid cell, (2) logged average population density, (3) presence of a major highway, and (4) logged distance to the closest major highway. All coefficients for the first three dependent variables are positive, indicating that route presence leads to higher levels of nighttime light intensity and population density, and a higher likelihood of major highway presence. Even in the ancient period, route presence is associated with a 35.8 percent increase in population density, and a 19.2 percentage point increase in the likelihood of a major highway being present in the cell. All coefficients for the distance to major highway variable are negative, indicating that grid cells with route presence in any period are generally closer to major highways than other cells. Presence of an ancient route is associated with a 54.4 percent increase in proximity to a major highway. All of the coefficients for route presence in the third period are of a greater magnitude than the corresponding coefficients for the second time period, and these are, in turn, higher than the magnitude of the corresponding coefficients from the ancient period. These results then imply that, while the routes of all periods are clearly still significant, the magnitude of these effects fades out over time.

In addition to testing the effect of route presence, I also test the effect of closer proximity to a route. Areas closer to the routes would have had more access to the benefits provided by them. Thus, if the Silk Routes did have a lasting impact, the cells closer to the routes ought to have higher levels of nighttime light intensity and population density compared to the cells further from the routes. The closest cells should also be more likely to have, and be proximate to, a highway. In order to test this relationship, I sort the cells into five 'bins' by distance from grid cell centroid to the closest route. This is done separately for each period, in order to analyze the relationship associated with proximity to a route *within* each time period. Then, I regress this new dummy variable on the same four dependent variables described above. I expect to find that the relationship between route presence and the dependent variables is highest in the bin with the cells closest to the route (those with centroids between 0-50 km away from the route), and progressively lower in the succeeding bins (50-100 km, 100-150 km, and 150-200 km, respectively).[50]

Table 6 shows the relationship between proximity to a route and the same dependent variables outlined above. In the Medieval period, the conditional association between route presence and the four dependent variables is significantly higher in cells that are 0-50 km away from the routes compared to those 50-100 km away from the routes. For example, while the effects of the route are significant in each distance bin, being within 0-50 km of a medieval route is associated with a 16.4 percent increase in nighttime light intensity compared to regions 200+ km away, while cells within 50-100 km of a medieval route are associated with a 3.9 percent increase in nighttime light intensity, cells within 100-150 km of a medieval route are associated with a 5.2 percent increase in nighttime light intensity, and cells within 150-200 km of a medieval route are associated with a 2.8 percent increase in nighttime light intensity (although this last result is not significant, the others are all significant at the 99 percent confidence level).

This relationship holds for each of the other two time periods as well — the areas closest to the routes generally have higher levels of nighttime light intensity and population density, and a higher likelihood of major highway presence. These locations are also closer to major highways. For all periods, the magnitude of the relationship decreases as cells get further away from the routes. The results shown in Table 6 demonstrate that even within these distance bins the relationship associated with the Mongol routes is consistently higher than the relationship

[50] Cells 200+ km away from the routes are used as the reference group.

associated with Tang-Caliphate period, which in turn is consistently higher than that associated with the Han-Roman period. This is true for every corresponding entry in the table.

In the ancient time period the bins closest to the routes (0-50 km) are the only areas to see these results highly significantly across the range of dependent variables, however in the second time period the results are significant in other bins for all variables except nighttime light intensity, and in the medieval period the results from other bins are significant for every dependent variable. This indicates that, in addition to the fade out over time that we see when comparing the magnitude of the coefficients across time periods and the fade out over space within each time period, the geographical scope of the effects also seems to consolidate over time — that is, the impact associated with ancient route presence seems to be concentrated in the areas immediately surrounding those locations whereas the impact for later periods is progressively more widely distributed around the routes. While it has been indicated that being located just outside of the sphere of influence of a transportation networks may have a *negative* spillover effect,[51] these baseline results do not clearly imply any such effect for the Silk Routes.

These results indicate that the effects of the routes fades over both time *and* space. The fact that the magnitude of the results is significantly different between each time period suggests that these relationships are due to the presence of the routes themselves rather than preferential geography — if geography was driving the original route placement as well as contemporary population density, economic activity and road placement, there would be no reason for the magnitude of the results to be different from route to route. Furthermore, if routes were chosen based on preferential location we would expect that the most advantageous terrain would have been selected in the earlier periods that had the most rudimentary technology. If this were true,

[51] Berger and Enflo, 2014.

and geography were the 'true' determinant, we would expect these regressions to have yielded an association that is the opposite of that seen in Tables 4-6.

Table 5: Route Presence and Core Variables of Interest

	(1) Ln(Nighttime Light Intensity)	(2) Ln(Population Density)	(3) Highway Presence	(4) Ln(Distance to Highway)
Medieval	0.173***	0.525***	0.245***	-0.749***
	(0.020)	(0.053)	(0.020)	(0.063)
Tang-	0.151***	0.508***	0.241***	-0.702***
Caliphate	(0.022)	(0.061)	(0.022)	(0.074)
Ancient	0.134***	0.358***	0.192***	-0.544***
	(0.025)	(0.061)	(0.024)	(0.081)
N	9,094	9,087	9,094	9,094

Notes: This table shows the relationship between route presence in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Each column denotes three separate regressions, one for each time period. The magnitude of the coefficients decreases in each older time period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Proximity to Route and Core Variables of Interest

	(1) Ln(Nighttime Light Intensity)	(2) Ln(Population Density)	(3) Highway Presence	(4) Ln(Distance to Highway)
		Medieval (ca. 1225-1335)		
0-50 km	0.164***	0.874***	0.228***	-1.009***
	(0.018)	(0.050)	(0.018)	(0.057)
50-100 km	0.039**	0.632***	0.078***	-0.607***
	(0.015)	(0.045)	(0.017)	(0.050)
100-150 km	0.052***	0.402***	0.087***	-0.455***
	(0.017)	(0.049)	(0.017)	(0.055)
150-200 km	0.028	0.215***	0.028	-0.249***
	(0.017)	(0.051)	(0.017)	(0.054)
		Tang-Caliphate Period (ca. 675-875)		
0-50 km	0.108***	0.787***	0.216***	-0.920***
	(0.019)	(0.057)	(0.020)	(0.064)
50-100 km	0.001	0.545***	0.068***	-0.539***
	(0.016)	(0.052)	(0.017)	(0.054)
100-150 km	0.019	0.255***	0.095***	-0.446***
	(0.018)	(0.055)	(0.018)	(0.060)
150-200 km	-0.021	0.059	0.041**	-0.190***
	(0.017)	(0.056)	(0.018)	(0.057)

Ancient Period (ca. 100 B.C. – A.D. 200)				
0-50 km	0.088***	0.460***	0.150***	-0.541***
	(0.021)	(0.055)	(0.022)	(0.068)
50-100 km	-0.021	0.263***	-0.007	-0.137**
	(0.018)	(0.054)	(0.018)	(0.054)
100-150 km	0.009	0.103*	0.035*	-0.122**
	(0.019)	(0.056)	(0.019)	(0.058)
150-200 km	-0.009	0.052	0.017	-0.051
	(0.018)	(0.057)	(0.019)	(0.057)
N	9,094	9,087	9,094	9,094

Notes: This table shows the relationship between proximity to a route in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For all time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases in each older time period, indicating that the relationship fades over both time and space. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Importance of Routes Away from Trading Centers

One major question that rises from the results outlined above is whether the significance of these results is due to the persistent importance of the city-state-like trading centers that served as the ‘nodes’ of the routes, rather than the paths of the routes themselves. To investigate whether the results of the previous section are being driven by the routes themselves, I conduct the same regressions as described in the previous section excluding grid cells that contain the ancient nodes.

The results of the presence regressions with nodes dropped out are shown in Table 7, and the corresponding distance regressions for the routes are shown in Table 8. The magnitude of the coefficients is slightly lower when the sites are dropped, however the same pattern of significance and persistence described in the previous section remain true for all regressions. Even in the ancient period, presence of a route is associated with 11.5 percent higher levels of nighttime light intensity and 32.8 percent higher levels of population density, as well as a 18.9 percentage point increase in the likelihood of having a major highway present and a 51.3 percent decrease in distance to a highway. The magnitude of the corresponding results in the Tang-Caliphate and Medieval periods grows progressively larger, again indicating that the persistence of the routes fades out over time. The fact that all of the results seem to be positive indicates that the Silk Roads were largely generating net positive change rather than simply relocating resources and wealth.

Table 7: Route Presence and Core Variables of Interest, Major Sites Dropped

	(1) Ln(Nighttime Light Intensity)	(2) Ln(Population Density)	(3) Highway Presence	(4) Ln(Distance to Highway)
Medieval	0.160^{***}	0.507^{***}	0.243^{***}	-0.731^{***}
	(0.019)	(0.053)	(0.020)	(0.062)
Tang-Caliphate	0.139^{***}	0.491^{***}	0.241^{***}	-0.683^{***}
	(0.021)	(0.062)	(0.023)	(0.073)
Ancient	0.115^{***}	0.328^{***}	0.189^{***}	-0.513^{***}
	(0.024)	(0.061)	(0.025)	(0.080)
N	9,078	9,071	9,078	9,078

Notes: This table shows the relationship between route presence in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways when the sites of major ancient trading centers are excluded from the regression. Each column denotes three separate regressions, one for each time period. The magnitude of the coefficients decreases in each older time period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Proximity to Route and Core Variables of Interest, Major Sites Dropped

	(1) Ln(Nighttime Light Intensity)	(2) Ln(Population Density)	(3) Highway Presence	(4) Ln(Distance to Highway)
	Medieval Period (ca. 1225-1335)			
0-50 km	0.154^{***}	0.863^{***}	0.226^{***}	-0.992^{***}
	(0.017)	(0.050)	(0.018)	(0.057)
50-100 km	0.038^{**}	0.631^{***}	0.078^{***}	-0.602^{***}
	(0.015)	(0.045)	(0.017)	(0.050)
100-150 km	0.052^{***}	0.403^{***}	0.087^{***}	-0.451^{***}
	(0.017)	(0.049)	(0.017)	(0.055)
150-200 km	0.028	0.216^{***}	0.027	-0.247^{***}
	(0.017)	(0.051)	(0.017)	(0.054)

Tang-Caliphate Period (ca. 675-875)				
0-50 km	0.099***	0.778***	0.216***	-0.903***
	(0.018)	(0.058)	(0.020)	(0.062)
50-100 km	0.000	0.546***	0.069***	-0.533***
	(0.016)	(0.052)	(0.017)	(0.054)
100-150 km	0.018	0.257***	0.095***	-0.441***
	(0.018)	(0.055)	(0.018)	(0.060)
150-200 km	-0.022	0.060	0.042**	-0.186***
	(0.017)	(0.056)	(0.018)	(0.057)
Ancient Period (ca 100 B.C. - A.D. 200)				
0-50 km	0.073***	0.441***	0.147***	-0.516***
	(0.020)	(0.055)	(0.022)	(0.067)
50-100 km	-0.022	0.264***	-0.006	-0.132**
	(0.018)	(0.054)	(0.018)	(0.054)
100-150 km	0.009	0.104*	0.035*	-0.117**
	(0.019)	(0.056)	(0.019)	(0.058)
150-200 km	-0.010	0.052	0.017	-0.047
	(0.018)	(0.057)	(0.019)	(0.057)
N	9,078	9,071	9,078	9,078

Notes: This table shows the relationship between proximity to a route in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways when the sites of major ancient trading centers are excluded from the regression. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For all time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases in each older time period, indicating that the relationship fades over both time and space. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In addition, to account for the possibility of ‘spheres of influence’ around these nodes, I conduct the same set of presence and distance regressions while dropping all grid cells with centroids within 150 km of the nodes identified above. The results of the presence regressions with the nodes and 150 km buffers dropped out are shown in Table 9, and the corresponding distance regressions for each time period are shown in Table 10.

The statistical significance does weaken slightly when both buffers and nodes are dropped from the analysis of the routes, however the same broader conclusions hold true with the effects both being significant and fading out over time and space. The magnitude of the coefficients generally drops as well, however even for the ancient period they still indicate a 7 percent increase in nighttime light intensity, a 28.3 percent increase in population density, and a 18.7 percentage point increase in the likelihood of having a road from the ancient period. Intriguingly, cutting out the buffers also increases the significance of the negative coefficients on both nighttime light intensity and road presence in the distance regressions for the ancient period, suggesting that perhaps there does exist a negative spillover effect that is somewhat obscured by the spheres of influence around cities in the baseline regressions.

Relative to cutting out only the ancient sites, cutting out the buffers restricts the measurement of presence to only the portions that meandered through the unregulated space between the oasis city-states. Given that the routes were unpaved, the magnitude and significance of these results is extraordinary. These results suggest that the paths ancient traders used to travel across Central Asia two millennia ago are still significantly impacting the economic activity and spatial distribution of the entire region today.

Table 9: Route Presence and Core Variables of Interest, 150 km Buffers Around Major Sites Dropped

	(1) Ln(Nighttime Light Intensity)	(2) Ln(Population Density)	(3) Highway Presence	(4) Ln(Distance to Highway)
Medieval	0.132***	0.506***	0.252***	-0.746***
	(0.019)	(0.059)	(0.023)	(0.068)
Tang-Caliphate	0.115***	0.514***	0.253***	-0.716***
	(0.023)	(0.070)	(0.025)	(0.080)
Ancient	0.070***	0.283***	0.187***	-0.470***
	(0.025)	(0.070)	(0.029)	(0.089)
N	8,736	8,729	8,736	8,736

Notes: This table shows the relationship between route presence in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways when cells with a centroid within 150 km of the site of a major ancient trading center are excluded from the regression. Each column denotes three separate regressions, one for each time period. The magnitude of the coefficients decreases in each older time period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Proximity to Route and Core Variables of Interest, 150 km Buffers Around Major Sites Dropped

	(1) Ln(Nighttime Light Intensity)	(2) Ln(Population Density)	(3) Highway Presence	(4) Ln(Distance to Highway)
		Medieval Period (ca. 1225-1335)		
0-50 km	0.116***	0.841***	0.223***	-0.964***
	(0.017)	(0.054)	(0.020)	(0.060)
50-100 km	0.017	0.630***	0.061***	-0.575***
	(0.014)	(0.049)	(0.017)	(0.053)
100-150 km	0.042**	0.400***	0.077***	-0.435***
	(0.017)	(0.051)	(0.018)	(0.056)
150-200 km	0.022	0.219***	0.024	-0.233***
	(0.017)	(0.052)	(0.017)	(0.054)
		Tang-Caliphate Period (ca. 675-875)		
0-50 km	0.080***	0.784***	0.218***	-0.897***
	(0.019)	(0.063)	(0.022)	(0.067)
50-100 km	-0.012	0.559***	0.047***	-0.514***
	(0.015)	(0.057)	(0.018)	(0.058)
100-150 km	0.005	0.254***	0.079***	-0.409***
	(0.017)	(0.057)	(0.019)	(0.061)
150-200 km	-0.027	0.068	0.036**	-0.167***
	(0.017)	(0.057)	(0.018)	(0.057)
		Ancient Period (ca. 100 B.C. – A.D. 200)		
0-50 km	0.018	0.368***	0.131***	-0.410***
	(0.020)	(0.062)	(0.024)	(0.072)
50-100 km	-0.059***	0.223***	-0.041**	-0.051
	(0.017)	(0.062)	(0.019)	(0.059)
100-150 km	-0.004	0.079	0.020	-0.083
	(0.020)	(0.061)	(0.019)	(0.061)
150-200 km	-0.016	0.053	0.012	-0.029
	(0.018)	(0.058)	(0.019)	(0.057)
N	8,736	8,729	8,736	8,736

Notes: This table shows the relationship between proximity to a route in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways when cells with a centroid within 150 km of the site of a major ancient trading center are excluded from the regression. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For all time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases in each older time period, indicating that the relationship fades over both time and space. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Potential Endogeneity of Geography

The degree to which geography may play a role in both the original route placements and contemporary economic distribution is the biggest potential weakness with these results — it is possible that the routes are in areas that are more naturally inclined to facilitate human life and activity and that the results associated with presence of a route are actually due to this preferential geography. This argument is weakened by the pattern of increasing significance through time that we see in the previous results. If geography were truly the motivating cause behind the significance of route placements, we would not expect to see different results for each time period because they ought to be equally affected by their respective geographies. Furthermore, the t-tests described in the data section indicate that there are no significant geographical differences between cells included in our data set that fall within the two 'bins' closest to the routes (i.e., between cells whose centroids were 0-50 km away from the routes vs. those that were 50-100 km away from the routes). However, it is still possible that I am not adequately controlling for the relevant geographical characteristics and that endogeneity related to preferential geographical locations is a problem.

To address this problem, I create a 'straight line instrument' based on straight lines drawn between the major nodes of the ancient routes, and use these placebo routes to predict route placement based on the shortest distance between two nodes connected by the road network. This method of using the most direct path between these major ancient trading centers (sites which any ancient route would have included) as an instrument will allow me to confirm that any relationship between route presence and the four outcomes measured can be attributed to the movement pattern caused by the Silk Road, rather than geographic determinants of the placement of the road.[52]

[52] Testing the geographical properties of the lines compared to the routes indicates that, where the straight lines diverge from the routes, the lines are more likely to be in desert, tundra or ice land cover categories. This indicates that the places where the routes diverge from the lines are instances where preferential geography might have played a role in route selection. Thus, the Two-Stage Least Squares regression is appropriately dropping out the cells that may have been chosen based on preferential geography.

The first stage results are shown in Table 11, and the 2SLS estimates using predicted routs are shown in Table 12 (for all time periods). All results are still significant, and the magnitude of the effects with this instrument are generally larger than the baseline results. The results again indicate a fade out of the effect over time, with presence of a route from the medieval period being associated with a 39.3 percent increase in nighttime light intensity, presence of a Tang-Caliphate period route being associated with a 37.7 percent increase in nighttime light intensity, and presence of an ancient route being associated with a 30.2 percent increase in nighttime light intensity.

Table 11: Straight Line Instrument First Stage

	(1) Medieval	(1) Tang-Caliphate	(1) Ancient
Straight Lines	0.170^{***}	0.165^{***}	0.274^{***}
Between Sites	(0.029)	(0.030)	(0.025)
F-statistic	33.300	30.159	123.522
N	8,149	8,149	11,163

Notes: This table shows the result of the first stage regression associated with the 'straight-line' instrument. The IV is designed around straight lines connecting the major ancient trading centers. The results indicate that there is a high conditional association between presence of a straight line and presence of a route in each periods. Each column denotes a separate regression. F-stats are all well above 10, indicating this is an appropriate instrument to use. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Straight Line Instrument 2SLS Regressions

	(1) Ln(Nighttime Light Intensity)	(2) Ln(Population Density)	(3) Highway Presence	(4) Ln(Distance to Highway)
Straight Lines	0.083^{***}	0.224^{***}	0.074^{***}	-0.303^{***}
Between Sites	(0.021)	(0.067)	(0.021)	(0.070)
Medieval	0.393^{***}	1.061^{***}	0.350^{***}	-1.437^{***}
	(0.104)	(0.330)	(0.095)	(0.330)
Tang-	0.377^{***}	1.017^{***}	0.335^{***}	-1.377^{***}
Caliphate	(0.099)	(0.318)	(0.092)	(0.322)
Ancient	0.302^{***}	0.817^{***}	0.269^{***}	-1.106^{***}
	(0.078)	(0.251)	(0.074)	(0.255)
N	11,160	11,153	11,160	11,160

Notes: This table shows the result of the Two-Stage Least Squares regression associated with the use of the straight-line instrument. Each column denotes four separate regressions, the first describing the relationship between presence of a straight line and the relevant dependent variable, the second describing the relationship between medieval route presence and the same dependent variable, the third describing the relationship between Tang-Caliphate route presence and the dependent variable, and the final entry describing the relationship between ancient route presence and the dependent variable. The results indicate that route presence in every period is still highly significant, with the consistently decreasing magnitude of coefficients from the medieval time period to the ancient time period exhibiting a fade out of the persistence over time. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Conclusion

In this thesis, I have shown that locations that had a Silk Route in any time period have higher contemporary levels of nighttime light intensity and population density, and are more proximate to major highways, relative to areas where the road did not exist. The magnitude of this relationship fades over both time and space. Furthermore, this relationship remains even when major nodes and their spheres of influence are excluded from the analysis, suggesting that at least some part of these outcomes is associated with presence of the routes themselves, rather than the highly-centralized cities along the routes. As the routes were generally little more than hiking paths that were traversed by people and livestock more often than they were by massive transportation infrastructures, these results indicate that even an institution as intangible as the Silk Roads can persist through thousands of years and continue to have a significant effect on modern economies and societies.

The Silk Road is on par with the Roman Empire in terms of ancient institutions that have so strong a cultural memory that they are universally recognizable today. This quantitative examination of such an influential institution has important implications not just for economic studies of trade and transportation, but also for historians studying trade networks and central Asia. Furthermore, while path dependence and economic activity are the two main variables of interest, this project also lays the groundwork for potential expansions that would allow for a more thorough examination of the effects of long-term international trading networks on the surrounding areas.

The analysis described in this thesis focuses purely on spatial dynamics and economic activity, while neglecting the considerable impact the Silk Road network had on political, technological and cultural aspects of both Europe and Asia. Future projects should exploit this varied impact to explore topics such as the effect of the Silk Road presence on contemporary religious, ethnic and linguistic diversity, or economic specialization. This current project taps only a fraction of the potential presented by the Silk Road, leaving open the opportunity for many future explorations of the impact of this monumental trade network.

More broadly, the strength of the results presented in this work suggest a fairly deterministic and long-lasting effect of the past upon the present. However, recent research has indicated that spatial equilibria can be shifted by certain shocks.[53] While the types of results found in this project are perhaps more immediately eye-catching, a deeper exploration of cases where ancient institutions are *not* deterministic will have important implications for contemporary economists and policy makers both in terms of understanding the nuances of the effects of the past upon the present and also in terms of implementing enduring policy actions.

[53] Redding, Sturm and Wulf.

Bibliography:

Acemoglu, D. & Robinson, J.A. (2012). *Why Nations Fail: The Origins of Power, Prosperity and Poverty. (*First ed). New York: Crown Publishers.

Barisitz, Stephan. (2017). *Central Asia and the Silk Road: Economic Rise and Decline over Several Millennia*. Springer, New York.

Berger, T., & Enflo, K. (2014). Locomotives of Local Growth: The Short- and Long-Term Impact of Railroads in Sweden. (Lund Papers in Economic History. General Issues; No. 132). Department of Economic History, Lund University.

Bleakley, H., & Lin, J. (2012). Portage and Path Dependence. *The Quarterly Journal of Economics*, 127(2), 587-644.

The Center for Geographic Analysis at Harvard. *Silk Road* [Data file]. Retrieved from: https://worldmap.harvard.edu/maps/7547.

Center for International Earth Science Information Network - CIESIN - Columbia University, And Information Technology Outreach Services - ITOS - University of Georgia. (2013).

Dalgaard, Carl-Johan & Kaarsen, Nicolai & Olsson, Ola & Selaya, Pablo, 2018. "Roman Roads to Prosperity: Persistence and Non-Persistence of Public Goods Provision," CEPR Discussion Papers 12745, C.E.P.R. Discussion Papers.

Datum: World Geodetic System 1984. *Georepository.* Retrieved from: https://georepository.com/datum_6326/World-Geodetic-System-1984.html

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica, 78*(6), 1863-1903.

Desmet, K., & Henderson, J.V. (2014). The Geography of Development within Countries. *CEPR Discussion Papers* DP10150.

Dickens, Andrew (2017). *Ethnolinguistic Favoritism in African Politics.*

Dignas, Beate and Engelbert Winter. *Rome and Persia in Late Antiquity: Neighbours and Rivals*. Cambridge: Cambridge University Press, 2007.

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5), 899-934.

Gallup, J. L., Sachs, J. D., & Mellinger, A. D. (1999). Geography and Economic Development. *International Regional Science Review*, *22*(2), 179–232. https://doi.org/10.1177/016001799761012334

Hansen, V. (2012). *The Silk Road: a new history.* Oxford University Press, New York.

J. Vernon Henderson, Adam Storeygard, and David N. Weil. (2012). "Measuring Economic Growth from Outer Space," *The American Economic Review* 102, no. 2: 994–1028.

IPUMS International. *World Map* [Data File]. Retrieved from: https://international.ipums.org/international/gis.shtml

Irwan, D. and Terviö, M. "Does Trade Raise Income?: Evidence from the Twentieth Century." *Journal of International Economics* 58(1): 1-18.

Jedwab, R., Kerby, E., & Moradi, A. (2017). History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya. *The Economic Journal*, 127(603), 1467-1494.

Martin, R., & Sunley, P. (2006). Path Dependence and regional economic evolution. *Journal of Economic Geography,* 6(4), 395-437.

Michaels, G., & Rauch, F. (2018). Resetting the Urban Network: 117-2012. *The Economic Journal,* 128(608), 378-412.

Michalopoulos, Stelios. (2012). "The Origins of Ethnolinguistic Diversity," *The American Economic Review* 102, no. 4: 1508–39.

MODIS Land Team. *Land Cover* [Data file]. Retrieved from: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MCD12C1_T1

Natural Earth. Coastline [Data file]. Retrieved from: https://www.naturalearthdata.com/downloads/10m-physical-vectors/

Natural Earth. Rivers + Lake Centerlines [Data file]. Retrieved from: https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines/.

Natural Earth. Roads [Data File]. Retrieved from: https://www.naturalearthdata.com/downloads/10m-cultural-vectors/roads/.

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

Shepherd, William R. and C.S. Hammond & Company. *Historical Atlas*. Eighth ed. Pikesville, Md:Colonial Offset Co, 1956.
Theil, Henri and Janet Galvez. “On Latitude and Affluence: The Equatorial Grand Canyon.” *Empirical Economics*, 20(1): 163-166.
United States Geological Survey. *Topography* [Data file]. Retrieved from: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MCD12C1_T1
Weidmann, Nils B., Jan Ketil Rød and Lars-Erik Cederman (2010). "Representing Ethnic Groups in Space: A New Dataset". Journal of Peace Research, in press.
Williams, T. (2014). *The Silk Roads: An ICOMOS Themeatic Study.* ICOMOS, Charenton-le Pont, France.
Whitfield, S. (2015). *Life along the Silk Road.* University of California Press, Oakland, Ca.
Zuleta, H. (2012). Seasonal fluctuations and economic growth. *Journal of Economic Development, 37(4),* 1-27.

Appendix

Section 1: Alternative Measures of Route Presence

As a robustness check for the route locations, I also created two more maps of the routes based on physical maps. One, the 'ancient routes' measure, is based on a map of trading networks, ca. 100 B.C.[54] The other, the 'medieval routes' measure, is based on a map from a historical atlas of trading networks during the 13th century.[55] These were both created by georectifying images of maps. I did not create a georectified version of the Tang-Caliphate period because I was unable to find a map that was appropriate in terms of both time period and geographical scope while also being spatially accurate enough to georectify (and thus match routes to specific coordinate locations).

Georectifying refers to the process of taking an image that is not currently attached to any known coordinate system and putting it into a known coordinate system. For each map, the process of georectification was roughly the same. First, scans of the map from each book were uploaded into ArcGIS. Then, I added ground control points (GCPs). GCPs are points for which the location is known in both the image as well as the target coordinate system. For the ancient map this consisted largely of clearly identifiable points such as river intersections and known site locations. For the historical atlas, many of the control points were at the intersections of latitude and longitudinal lines because these lines were included in the image itself.

ArcGIS supports affine, similarity and projective transformations. ArcGIS requires a minimum of three control points for a first order (affine) transformation, six for a second order and

[54] Dignas, Beate and Engelbert Winter. *Rome and Persia in Late Antiquity: Neighbours and Rivals*. Cambridge: Cambridge University Press, 2007.
[55] Shepherd, William R. and C.S. Hammond & Company. *Historical Atlas*. Eighth ed. Pikesville, Md: Colonial Offset Co, 1956.

nine for a third order. All transformations in this paper were done using a third order polynomial (to maximize the accuracy of the transformation), built on the control points and using a least squares fitting (LSF) algorithm. At least 25 control points were used for each transformation, well exceeding the required minimum. The rectified images are shown below in Figures 3 and 4, with the outlines of modern coastline locations shown in blue. Figures 3a and 4a depict the original scans of the maps, while figures 3b and 4b depict the georectified versions.

Figure 3a: Original Photos of Ancient Routes

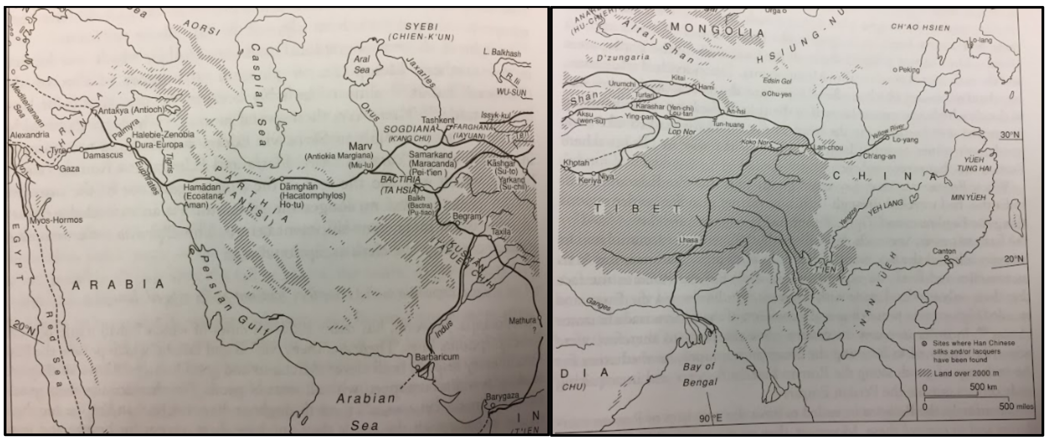


Figure 3b: Transformed Photos of Ancient Routes

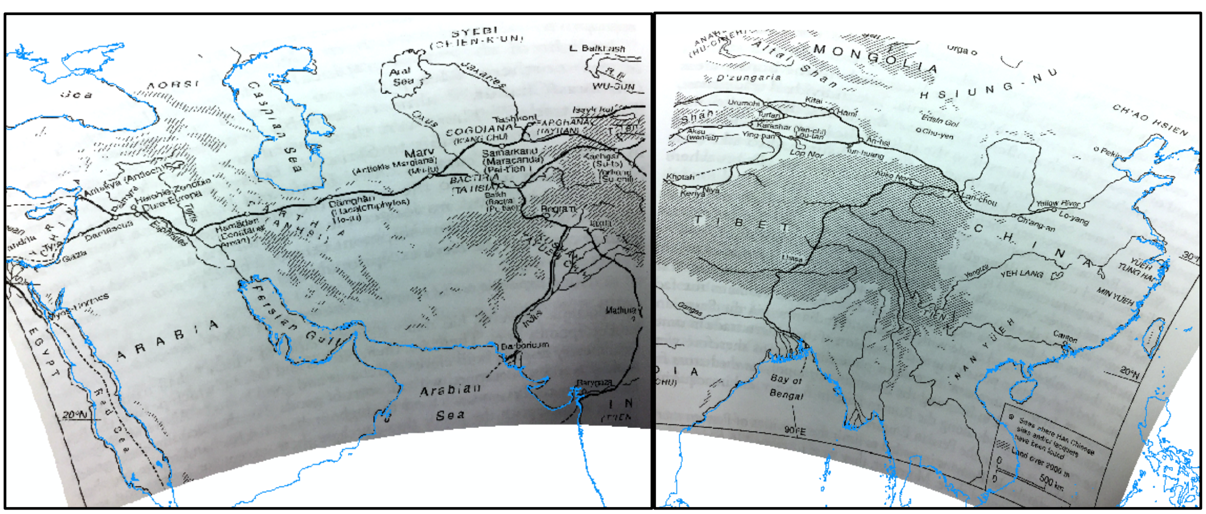


Figure 4a: Original Photo of Medieval Routes

Figure 4b: Transformed Photo of Medieval Routes

After transforming the images, ArcGIS was used to trace the routes from the rectified image of the map within a new shapefile. The result is a new map file for each period, with the paper maps having been completely digitized into a shapefile map of the routes. Figures 5 and 6 below show the resultant shapefile of each time period, overlaid onto a base map of the world. This shapefile was then combined with the grid cell data using the 'Spatial Join' tool as described above. Out of the 11,456 grid cells, 95.32% show no ancient route presence and are coded as 0. Of the remaining 536, 115 grid cells contain part of more than one ancient route. All cells with any form of route presence have been coded as 1. Out of the 11,456 grid cells, 95.32% show no route presence and are coded as 0. Of the remaining 536, 115 grid cells contain part of more than one route. All cells with any form of route presence have been coded as 1.

Figure 5: Shapefile of Ancient Routes

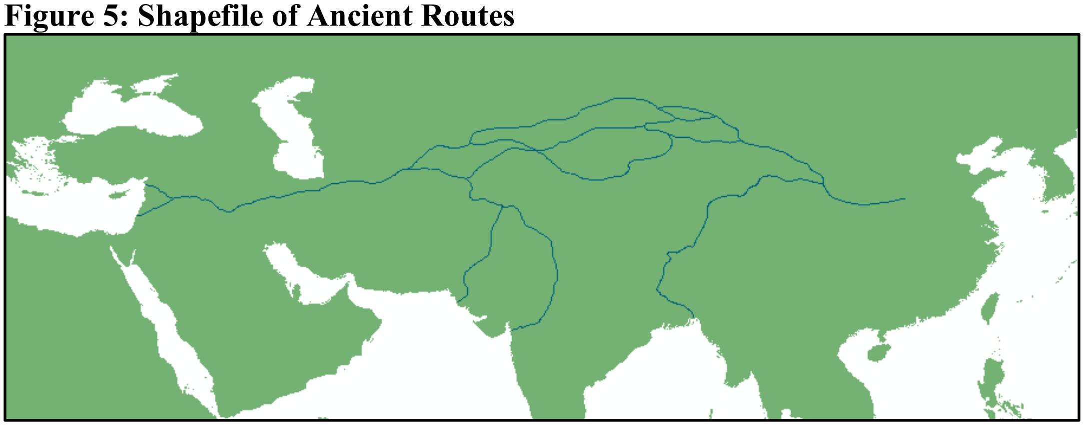
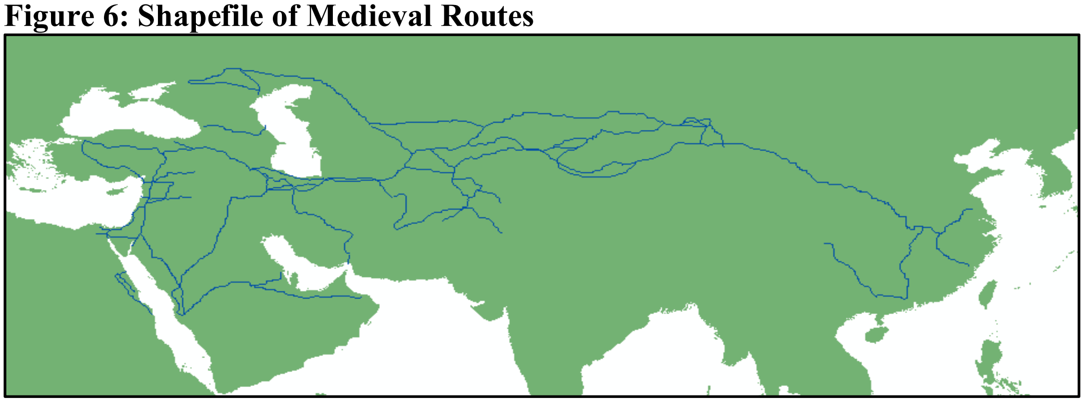


Figure 6: Shapefile of Medieval Routes



Results[56]

Table 1.1: Medieval Geography T-tests

	(1) 0-50 km	(2) 50-100 km	(3) Diff.	(4) Std. Error	(5) Obs.
Elevation	99.0610	102.5832	-3.5223	2.3350	1,780
SD Elevation	10.6466	10.0041	0.6425	0.5199	1,802
Dist. to River	94.45	98.501	-4.051	4.021	1,877
Dist. to Coast	947.664	940.377	7.287	33.349	1,877
Forest (LC)	0.0233	0.0262	-0.0029	0.0072	1,877
Shrubland (LC)	0.1330	0.1359	-0.0029	0.0159	1,877
Grasslands (LC)	0.3647	0.3766	-0.0119	0.0225	1,877
Barren (LC	0.3274	0.3329	-0.0055	0.0219	1,877
Forest (PV)	0.1033	0.0960	0.0072	0.0140	1,877
Shrubland (PV)	0.3433	0.3354	0.0078	0.0221	1,877
Grasslands (PV)	0.0540	0.0374	0.0165*	0.0099	1,877
Tundra/Ice (PV)	0.2400	0.2282	0.0118	0.0198	1,877
Desert (PV)	0.0009	0.0262	-0.0253***	0.0050	1,877

Notes: I use two separate measures of landcover/climate: contemporary land cover, and potential vegetation. LC denotes a contemporary land cover category, and PV denotes a potential vegetation category. Column 1 indicates the average for cells with a centroid within 0-50 km of the closest medieval (Mongol) route, and column 2 indicates the average for cells with a centroid within 50-100 km of the closest medieval route. Column 3 indicates the difference between the averages in columns 1 and 2, column 4 indicates the standard error, and column 5 indicates the number of observations in the t-test. * p<0.10, ** p<0.05, *** p<0.01.

Table 1.2: Ancient Geography T-tests

	(1) 0-50 km	(2) 50-100 km	(3) Diff.	(4) Std. Error	(5) Obs.
Elevation	96.3108	104.0244	-7.7136***	2.9442	1,096
SD Elevation	9.7551	9.7989	-0.0438	0.6811	1,097
Dist. to River	83.755	87.460	-3.705	4.997	1,097
Dist. to Coastline	1,070.309	1,041.488	28.821	38.186	1,097
Forest (LC)	0.0195	0.0393	-0.0198*	0.0102	1,097
Shrubland (LC)	0.1137	0.1742	-0.0605***	0.0211	1,097
Grasslands (LC)	0.1705	0.1948	-0.0242	0.0233	1,097
Barren (LC	0.3925	0.3558	0.0367	0.0292	1,097
Forest (PV)	0.1439	0.1610	-0.0172	0.0217	1,097
Shrubland (PV)	0.3730	0.3258	0.0472	0.0288	1,097
Grasslands (PV)	0.0746	0.0768	-0.0022	0.0160	1,097
Tundra/Ice (PV)	0.2913	0.2753	0.0160	0.0272	1,097
Desert (PV)	0.0089	0.0337	-0.0248***	0.0086	1,097

Notes: I use two separate measures of landcover/climate: contemporary land cover, and potential vegetation. LC denotes a contemporary land cover category, and PV denotes a potential vegetation category. Column 1 indicates the average for cells with a centroid within 0-50 km of the closest ancient route, and column 2 indicates the average for cells with a centroid within 50-100 km of the closest ancient route. Column 3 indicates the difference between the averages in columns 1 and 2, column 4 indicates the standard error, and column 5 indicates the number of observations in the t-test. * p<0.10, ** p<0.05, *** p<0.01.

[56] All regressions are run using potential vegetation as the measure of landcover, and with 'desert' cells dropped.

Table 1.3: Route Presence and Nighttime Light Intensity

	Dependent Variable: Nighttime Light Intensity					
	(1)	(2)	(3)	(4)	(5)	(6)
Medieval	0.068***	0.124***	0.120***	0.114***	0.120***	0.096***
	(0.019)	(0.019)	(0.018)	(0.017)	(0.020)	(0.016)
Ancient	0.180***	0.141***	0.228***	0.135***	0.150***	0.099***
	(0.030)	(0.027)	(0.026)	(0.024)	(0.031)	(0.023)
Political Controls		x		x	x	x
Geo. Controls			x	x	x	x
Potential Vegetation					x	x
Landcover				x		
Desert Dropped						x
N	11,456	11,456	11,282	11,282	9,094	11,282

Notes: Each column in this table indicates two separate regressions: one using a dummy for medieval route presence, and one using a dummy for ancient route presence. Column 1 indicates the results of a simple OLS regression of route presence on nighttime light intensity. Column 2 indicates the results of a regression including only the vector of political controls (with fixed effects for the country and ethnicities located at the grid cell centroid) in addition to route presence, while Column 3 indicates the results of a regression including only the vector of geographical controls (with fixed effects for potential vegetation) in addition to route presence. Column 4 includes the results of the regression of route presence including both political and geographical controls but using contemporary landcover instead of potential vegetation, while Column 5 indicates the results of the regression using both sets of controls with potential vegetation rather than contemporary landcover. Finally, column 6 indicates the results of the same regression in column 5 but with desert cells dropped out. This is the preferred specification used for the rest of the paper. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.4: Route Presence and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
Medieval	0.102***	0.422***	0.146***	-0.473***
	(0.019)	(0.054)	(0.020)	(0.059)
Ancient	0.150***	0.358***	0.119***	-0.377***
	(0.031)	(0.064)	(0.029)	(0.087)
N	9,094	9,087	9,094	9,094

Notes: This table shows the relationship between route presence in each period and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Each column denotes two separate regressions, one for each time period. The magnitude of the coefficients decreases from the medieval period to the ancient period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.5: Proximity to Route and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
		Medieval (ca. 1225-1335)		
0-50 km	0.066***	0.763***	0.177***	-0.863***
	(0.018)	(0.053)	(0.019)	(0.056)
50-100 km	-0.027*	0.574***	0.100***	-0.706***
	(0.016)	(0.054)	(0.019)	(0.061)
100-150 km	-0.041***	0.377***	0.071***	-0.493***
	(0.015)	(0.054)	(0.017)	(0.053)
150-200 km	-0.060***	0.090	0.030*	-0.353***
	(0.015)	(0.057)	(0.018)	(0.062)
		Ancient Period (ca. 100 B.C. – A.D. 200)		
0-50 km	0.163***	0.632***	0.120***	-0.552***
	(0.026)	(0.061)	(0.026)	(0.074)
50-100 km	0.074***	0.517***	0.020	-0.352***
	(0.022)	(0.054)	(0.023)	(0.066)
100-150 km	0.061***	0.386***	0.060***	-0.359***
	(0.021)	(0.059)	(0.022)	(0.066)
150-200 km	0.018	0.256***	0.062***	-0.310***
	(0.017)	(0.059)	(0.023)	(0.063)
N	9,094	9,087	9,094	9,094

Notes: This table shows the relationship between proximity to a route in each period and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For both time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases from medieval to ancient, indicating that the relationship fades over both time and space. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Section 2: Including Desert Cells[57]

Table 2.1: Route Presence and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
Medieval	0.165***	0.431***	0.241***	-0.677***
	(0.017)	(0.052)	(0.018)	(0.057)
Tang-Caliphate	0.148***	0.394***	0.234***	-0.626***
	(0.018)	(0.058)	(0.019)	(0.063)
Ancient	0.131***	0.236***	0.174***	-0.427***
	(0.020)	(0.061)	(0.020)	(0.068)
N	11,282	11,275	11,282	11,282

Notes: This table shows the relationship between route presence in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Each column denotes three separate regressions, one for each time period. The magnitude of the coefficients decreases in each older time period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.2: Proximity to Route and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
		Medieval (ca. 1225-1335)		
0-50 km	0.155***	0.752***	0.211***	-0.903***
	(0.016)	(0.049)	(0.016)	(0.051)
50-100 km	0.036***	0.564***	0.054***	-0.546***
	(0.013)	(0.045)	(0.014)	(0.044)
100-150 km	0.050***	0.436***	0.073***	-0.438***
	(0.015)	(0.047)	(0.015)	(0.047)
150-200 km	0.024	0.235***	0.023	-0.227***
	(0.015)	(0.049)	(0.014)	(0.046)

[57] All of these regressions include both vectors of controls, with potential vegetation as the measure of land cover. Desert cells are included in the regressions.

Tang-Caliphate Period (ca. 675-875)				
0-50 km	0.104***	0.654***	0.198***	-0.808***
	(0.016)	(0.054)	(0.017)	(0.055)
50-100 km	-0.006	0.490***	0.043***	-0.471***
	(0.013)	(0.050)	(0.014)	(0.046)
100-150 km	0.007	0.325***	0.076***	-0.408***
	(0.015)	(0.050)	(0.015)	(0.049)
150-200 km	-0.027*	0.087*	0.031**	-0.162***
	(0.015)	(0.052)	(0.014)	(0.047)
Ancient Period (ca. 100 B.C. – A.D. 200)				
0-50 km	0.084***	0.304***	0.118***	-0.371***
	(0.017)	(0.055)	(0.018)	(0.058)
50-100 km	-0.021	0.186***	-0.030**	-0.047
	(0.015)	(0.053)	(0.015)	(0.045)
100-150 km	-0.011	0.120**	0.020	-0.095*
	(0.016)	(0.051)	(0.015)	(0.048)
150-200 km	-0.026*	0.045	0.007	-0.030
	(0.015)	(0.052)	(0.016)	(0.048)
N	11,282	11,275	11,282	11,282

Notes: This table shows the relationship between proximity to a route in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For all time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases in each older time period, indicating that the relationship fades over both time and space. I use the full set of controls. Robust SE in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Section 3: Potential Vegetation Alternate: Contemporary Landcover[58]

Table 3.1: Route Presence and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
Medieval	0.109^{***}	0.406^{***}	0.212^{***}	-0.505^{***}
	(0.021)	(0.078)	(0.028)	(0.076)
Tang-Caliphate	0.078^{***}	0.345^{***}	0.193^{***}	-0.424^{***}
	(0.022)	(0.089)	(0.030)	(0.083)
Ancient	0.061^{**}	0.137	0.115^{***}	-0.165^{*}
	(0.024)	(0.093)	(0.031)	(0.089)
N	2,806	2,806	2,806	2,806

Notes: This table shows the relationship between route presence in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Each column denotes three separate regressions, one for each time period. The magnitude of the coefficients decreases in each older time period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.2: Proximity to Route and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
		Medieval (ca. 1225-1335)		
0-50 km	0.136^{***}	0.701^{***}	0.273^{***}	-0.846^{***}
	(0.027)	(0.094)	(0.033)	(0.095)
50-100 km	0.015	0.403^{***}	0.141^{***}	-0.457^{***}
	(0.023)	(0.085)	(0.032)	(0.079)
100-150 km	0.024	0.524^{***}	0.110^{***}	-0.494^{***}
	(0.024)	(0.092)	(0.035)	(0.098)
150-200 km	-0.002	0.218^{**}	0.017	-0.222^{***}
	(0.022)	(0.092)	(0.030)	(0.080)

[58] All of these regressions include both vectors of controls, with contemporary land cover as the measure of landcover. No categories of contemporary land cover are excluded.

Tang-Caliphate Period (ca. 675-875)				
0-50 km	0.085***	0.606***	0.257***	-0.749***
	(0.027)	(0.106)	(0.035)	(0.103)
50-100 km	-0.019	0.337***	0.124***	-0.397***
	(0.023)	(0.093)	(0.033)	(0.083)
100-150 km	-0.023	0.430***	0.095***	-0.404***
	(0.022)	(0.094)	(0.034)	(0.096)
150-200 km	-0.029	0.069	0.011	-0.134
	(0.022)	(0.101)	(0.031)	(0.083)
Ancient Period (ca. 100 B.C. – A.D. 200)				
0-50 km	0.078***	0.284**	0.139***	-0.356***
	(0.030)	(0.110)	(0.036)	(0.107)
50-100 km	-0.038*	0.060	0.000	0.011
	(0.022)	(0.100)	(0.034)	(0.087)
100-150 km	-0.039	0.157	-0.011	-0.151
	(0.025)	(0.096)	(0.032)	(0.104)
150-200 km	-0.029	-0.005	-0.010	-0.016
	(0.022)	(0.100)	(0.031)	(0.080)
N	2,806	2,806	2,806	2,806

Notes: This table shows the relationship between proximity to a route in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For all time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases in each older time period, indicating that the relationship fades over both time and space. I use the full set of controls. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Section 4: Clustered Standard Errors

Table 4.1: Route Presence and Nighttime Light Intensity

	Dependent Variable: Nighttime Light Intensity					
	(1)	(2)	(3)	(4)	(5)	(6)
Medieval (ca. 1225 -1335)	0.153***	0.159***	0.203***	0.135***	0.165***	0.175***
	(0.054)	(0.037)	(0.044)	(0.033)	(0.033)	(0.038)
Tang-Caliphate (ca. 675-875)	0.131*	0.149***	0.185***	0.115***	0.148***	0.155***
	(0.075)	(0.052)	(0.045)	(0.036)	(0.039)	(0.037)
Ancient (ca. 100 B.C. – 200 A.D.)	0.124**	0.144***	0.149***	0.092**	0.131***	0.134***
	(0.052)	(0.047)	(0.048)	(0.034)	(0.038)	(0.047)
Political Controls		x		x	x	x
Geo. Controls			x	x	x	x
Potential Vegetation					x	x
Landcover				x		
Desert Dropped						x
N	11,456	11,456	11,282	11,282	11,282	9,094

Notes: Each column in this table indicates three separate regressions: one using a dummy for medieval route presence, one using a dummy for route presence in the Tang-Caliphate period, and one using a dummy for ancient route presence. Column 1 indicates the results of a simple OLS regression of route presence on nighttime light intensity. Column 2 indicates the results of a regression including only the vector of political controls (with fixed effects for the country and ethnicities located at the grid cell centroid) in addition to route presence, while Column 3 indicates the results of a regression including only the vector of geographical controls (with fixed effects for potential vegetation) in addition to route presence. Column 4 includes the results of the regression of route presence including both political and geographical controls but using contemporary landcover instead of potential vegetation, while Column 5 indicates the results of the regression using both sets of controls with potential vegetation rather than contemporary landcover. Finally, column 6 indicates the results of the same regression in column 5 but with desert cells dropped out. This is the preferred specification used for the rest of the paper. Clustered standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.2: Route Presence and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
Medieval	0.173***	0.525***	0.245***	-0.749***
	(0.035)	(0.111)	(0.051)	(0.167)
Tang-Caliphate	0.151***	0.508***	0.241***	-0.702***
	(0.037)	(0.094)	(0.055)	(0.171)
Ancient	0.134***	0.358***	0.192***	-0.544***
	(0.047)	(0.117)	(0.052)	(0.186)
N	9,094	9,087	9,094	9,094

Notes: This table shows the relationship between route presence in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Each column denotes three separate regressions, one for each time period. The magnitude of the coefficients decreases in each older time period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls, with desert cells excluded. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.3: Proximity to Route and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
		Medieval (ca. 1225-1335)		
0-50 km	0.164**	0.874***	0.228***	-1.009***
	(0.077)	(0.233)	(0.047)	(0.218)
50-100 km	0.039	0.632***	0.078***	-0.607***
	(0.071)	(0.221)	(0.021)	(0.146)
100-150 km	0.052	0.402**	0.087***	-0.455***
	(0.061)	(0.194)	(0.031)	(0.135)
150-200 km	0.028	0.215	0.028*	-0.249***
	(0.048)	(0.133)	(0.015)	(0.044)
		Tang-Caliphate Period (ca. 675-875)		
0-50 km	0.108	0.787***	0.216***	-0.920***
	(0.065)	(0.220)	(0.046)	(0.228)
50-100 km	0.001	0.545**	0.068***	-0.539***
	(0.060)	(0.229)	(0.023)	(0.168)
100-150 km	0.019	0.255	0.095***	-0.446***
	(0.047)	(0.186)	(0.029)	(0.149)
150-200 km	-0.021	0.059	0.041**	-0.190***
	(0.037)	(0.127)	(0.018)	(0.056)
		Ancient Period (ca. 100 B.C. – A.D. 200)		
0-50 km	0.088	0.460*	0.150***	-0.541*
	(0.075)	(0.226)	(0.054)	(0.268)
50-100 km	-0.021	0.263	-0.007	-0.137
	(0.067)	(0.242)	(0.022)	(0.174)
100-150 km	0.009	0.103	0.035	-0.122
	(0.055)	(0.217)	(0.042)	(0.194)
150-200 km	-0.009	0.052	0.017	-0.051
	(0.050)	(0.183)	(0.020)	(0.116)
N	9,094	9,087	9,094	9,094

Notes: This table shows the relationship between proximity to a route in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For all time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases in each older time period, indicating that the relationship fades over both time and space. I use the full set of controls, with desert cells excluded. Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Section 5: 1x1 Grid Cells

Table 5.1: Route Presence and Nighttime Light Intensity

	Dependent Variable: Nighttime Light Intensity					
	(1)	(2)	(3)	(4)	(5)	(6)
Medieval	0.087^{***}	0.093^{***}	0.156^{***}	0.117^{***}	0.144^{***}	0.109^{***}
(ca. 1225 -1335)	(0.026)	(0.026)	(0.023)	(0.022)	(0.025)	(0.021)
Tang-Caliphate	0.056^{**}	0.080^{***}	0.131^{***}	0.088^{***}	0.101^{***}	0.078^{***}
(ca. 675-875)	(0.026)	(0.027)	(0.025)	(0.023)	(0.027)	(0.022)
Ancient	0.054^{*}	0.089^{***}	0.082^{***}	0.072^{***}	0.082^{***}	0.061^{**}
(ca. 100 B.C. – 200 A.D.)	(0.030)	(0.029)	(0.027)	(0.025)	(0.031)	(0.024)
Political Controls		x		x	x	x
Geo. Controls			x	x	x	x
Potential Vegetation					x	x
Landcover				x		
Desert Dropped						x
N	2,849	2,849	2,806	2,806	2,251	2,806

Notes: Each column in this table indicates three separate regressions: one using a dummy for medieval route presence, one using a dummy for route presence in the Tang-Caliphate period, and one using a dummy for ancient route presence. Column 1 indicates the results of a simple OLS regression of route presence on nighttime light intensity. Column 2 indicates the results of a regression including only the vector of political controls (with fixed effects for the country and ethnicities located at the grid cell centroid) in addition to route presence, while Column 3 indicates the results of a regression including only the vector of geographical controls (with fixed effects for potential vegetation) in addition to route presence. Column 4 includes the results of the regression of route presence including both political and geographical controls but using contemporary landcover instead of potential vegetation, while Column 5 indicates the results of the regression using both sets of controls with potential vegetation rather than contemporary landcover. Finally, column 6 indicates the results of the same regression in column 5 but with desert cells dropped out. This is the preferred specification used for the rest of the paper. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.2: Route Presence and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
Medieval	0.133^{***}	0.550^{***}	0.244^{***}	-0.608^{***}
	(0.025)	(0.080)	(0.032)	(0.092)
Tang-	0.101^{***}	0.516^{***}	0.236^{***}	-0.519^{***}
Caliphate	(0.027)	(0.096)	(0.036)	(0.107)
Ancient	0.082^{***}	0.348^{***}	0.191^{***}	-0.354^{***}
	(0.031)	(0.092)	(0.036)	(0.111)
N	2,251	2,251	2,251	2,251

Notes: This table shows the relationship between route presence in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Each column denotes three separate regressions, one for each time period. The magnitude of the coefficients decreases in each older time period, although maintaining statistical significance, indicating that the relationship fades over time. I use the full set of controls with desert cells removed. Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.3: Proximity to Route and Core Variables of Interest

	(1) Nighttime Light Intensity	(2) Population Density	(3) Highway Presence	(4) Distance to Highway
	Medieval (ca. 1225-1335)			
0-50 km	0.176***	0.883***	0.329***	-0.997***
	(0.030)	(0.096)	(0.037)	(0.114)
50-100 km	0.040	0.459***	0.172***	-0.547***
	(0.026)	(0.083)	(0.037)	(0.094)
100-150 km	0.058*	0.551***	0.170***	-0.615***
	(0.030)	(0.101)	(0.043)	(0.129)
150-200 km	0.008	0.181*	0.044	-0.263***
	(0.024)	(0.095)	(0.038)	(0.101)
	Tang-Caliphate Period (ca. 675-875)			
0-50 km	0.118***	0.774***	0.320***	-0.871***
	(0.031)	(0.113)	(0.041)	(0.130)
50-100 km	0.006	0.358***	0.157***	-0.484***
	(0.028)	(0.095)	(0.040)	(0.105)
100-150 km	0.004	0.370***	0.156***	-0.500***
	(0.028)	(0.104)	(0.043)	(0.127)
150-200 km	-0.031	-0.066	0.041	-0.125
	(0.028)	(0.113)	(0.041)	(0.111)
	Ancient Period (ca. 100 B.C. – A.D. 200)			
0-50 km	0.109***	0.109***	0.530***	0.223***
	(0.037)	(0.037)	(0.108)	(0.042)
50-100 km	-0.026	-0.026	0.060	0.044
	(0.028)	(0.028)	(0.097)	(0.041)
100-150 km	-0.019	-0.019	0.120	0.014
	(0.033)	(0.033)	(0.105)	(0.040)
150-200 km	-0.028	-0.028	-0.070	0.018
	(0.028)	(0.028)	(0.114)	(0.039)
N	2,251	2,251	2,251	2,251

Notes: This table shows the relationship between proximity to a route in each of the three periods and contemporary levels of nighttime light intensity, population density, likelihood of highway presence and proximity to highways. Grid cells are sorted into five bins, with the first including grid cells with a centroid within 0-50 km of the route, the second including cells with a centroid within 50-100 km of the route, the third including cells with a centroid within 100-150 km of the route, the fourth including cells with a centroid within 150-200 km of the route, and the last including cells with a centroid 200+ km away from the route. This last category is used as the reference group in the regression. For all time periods the magnitude of the coefficients for cells within 0-50 km of the route is larger than those for cells within 50-100 km of the route, indicating that the relationship between route presence and the four measures of persistence fades out as distance to the route increases. Each time period is a separate regression. Within each distance bin, the magnitude of the coefficients still decreases in each older time period, indicating that the relationship fades over both time and space. I use the full set of controls with desert cells removed. Robust SE in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.