UNIVERSITY OF HOHENHEIM

Hohenheim Discussion Papers in Business, Economics and Social Sciences

GO EAST: ON THE IMPACT OF THE TRANSIBERIAN RAILWAY ON ECONOMIC DEVELOPMENT IN EASTERN RUSSIA

Sebastian Seiffert
University of Hohenheim

Research Area INEPA

02-2019

wiso.uni-hohenheim.de

Discussion Paper 02-2019

GO EAST: ON THE IMPACT OF THE TRANSIBERIAN RAILWAY ON ECONOMIC DEVELOPMENT IN EASTERN RUSSIA

Sebastian Seiffert

Research Area "INEPA – Inequality and Economic Policy Analysis"

Download this Discussion Paper from our homepage:

https://wiso.uni-hohenheim.de/papers

ISSN 2364-2084

Die Hohenheim Discussion Papers in Business, Economics and Social Sciences dienen der schnellen Verbreitung von Forschungsarbeiten der Fakultät Wirtschafts- und Sozialwissenschaften. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung der Fakultät Wirtschafts- und Sozialwissenschaften dar.

Hohenheim Discussion Papers in Business, Economics and Social Sciences are intended to make results of the Faculty of Business, Economics and Social Sciences research available to the public in order to encourage scientific discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the Faculty of Business, Economics and Social Sciences.

GO EAST:
ON THE IMPACT OF THE TRANSIBERIAN RAILWAY ON ECONOMIC DEVELOPMENT IN EASTERN RUSSIA

*

Sebastian Seiffert†

January 24, 2019

Abstract

This paper addresses the question whether or not large-scale infrastructure investments have a causal effect of local economic development. By using a novel instrumental variable approach based on historical trade and travel routes across the Russian East, I am able to identify a causal and negative effect of remoteness to the Transsiberian Railway on local economic activity as measured by nocturnal lights emission.

Keywords: Transport Costs, Railway, Russia, Nightlights, Regional Economics, Development
JEL-Classifications: O11, O18, R11, R40.

*I would like to thank Thomas Beißinger for his guidance throughout this project. Further, my thanks go to Sibylle Lehmann-Hasenmeyer, Fabian Wahl and the participants of the seminar in Economic History at the University of Hohenheim. All errors are my own.

†University of Hohenheim, Department of Economics, Schloss Hohenheim, 70593 Stuttgart, Germany, phone +49 (0) 711 459 23245, email: seiffert@uni-hohenheim.de.

1 Introduction

A sizeable amount of the annual World Bank budget is being allocated towards transport infrastructure improvements year by year. High trade costs are seen as key impediment to local economic development. It has been shown by a large range of theoretical contributions that reductions in trade costs result in higher levels of real income in trading as well as alleviate the impact of shocks (Donaldson, 2018). But can such investments in infrastructure be seen as the go-to method in order to foster local development? Is it not likely that also contexts, i.e. the specific local circumstances such as existing spatial equilibria or local natural endowments also matter? It is much likely that the impact of local measures is affected by the level of local attributes such as resource availability or the suitability for agriculture. While we have a good understanding of the theoretical mechanisms behind the nexus of trade costs and economic development, the empirical literature to this day remains quite sparse.

The historically most prominent and wide-spread strategy in order to reduce transport costs is the construction of railroad networks. It has been widely used in both the transformation of western countries towards industrialised economies as well as in the annexation and exploitation of countries in the era of colonialism. This opens the question if the spatial distribution of economic activity in such contexts is entirely determined by local natural advantages or if path dependence plays a decisive role also in the long run. The main challenge in order to empirically assess the existence of long run path dependence is to separate the effects of state dependence which is the local availability of factors of production from serial correlation which can generally be described as advantages which first attracted other factors (Bleakley and Lin, 2012). One of those additional factors which clearly matter are local differences in trade costs and with it differences in the accessibility of remote markets.

In the underlying paper, I investigate the long-run causal effect of the construction of the Trans-Siberian Railway (henceforth: TSR) and the implied reduction in local transport costs on the level and spatial organisation of economic activity in Asian Russia. The assessment is based on data nocturnal lights emission, the historical as well as current network of railways in Eastern Russia as well as a wide array of control variables.

In order empirically disentangle the effects of local natural endowments from the effect of the TSR, I suggest to solve this endogeneity problem by employing data on historical tea trade in Tsarist Russia. The so-called Tea Road connected European Russia with first Nerchensk in Manchuria and later Kyaktha located on what is now the border between Russia and Mongolia. As there were hardly any local markets east of the Ural mountains at the time and the fact that the route was not suitable for the transport of large amounts of heavy goods as well as persons, the route was a mere transit route. Its' main purpose was to link the Chinese tea production sites with the tea markets in European Russia. This route was later extended to Vladivostok and used as a Post Route in addition to its original purpose. While the area which today constitutes the districts of Ural (in parts), Siberia and the Far East came under Russian control as early as the late 17th century, the population remained below 300.000 until the early 20th century. The construction of the TSR set in motion a large-scale settling process which lead to an increase in the population up to about 36 million in current times[1]. Accordingly, upon its completion the TSR facilitated both the colonisation of Eastern Russia as well as created a strong focal line for economic activity due to an early change in transport costs along its' course. Hence, Asian Russia offers a research setting which - to my knowledge - is unique. Using this historical trade route as an instrument, I show that being located further away from the TSR has a causal and negative impact on contemporary economic activity as recorded by nocturnal lights emission. I further create a new data set linking information on railroads, historical (trade) routes, economic activity and local natural endowments at a high spatial granularity of 132,843 grid cells with the size of 0.1x0.1 decimal degrees which is equivalent to 11x11km.

I show how the construction of the TSR and the associated reduction in transport costs facilitated the colonisation of the Russian East. Until its construction the area it now transects was very scarcely populated and did not exhibit any noteworthy economic activity. Within the following 100 years the population in its vicinity highly increased and today there exists sizeable economic activity. By providing a novel instrumental variable estimation framework, I am able to demonstrate the causal effects of the reduction in transport costs.

In doing so, I contribute to the ongoing scientific discussion about the causal impact of large-scale infrastructure projects. Namely, to answering the question whether a large-scale transport infrastructure project and the implied reduction in trade costs have a causal impact on local economic growth. This is possible since with the underlying setup I am able disentangle the effects of such an investment from the ones of local endowments in a context of non-existing scale effects. Choosing the TSR is mainly driven by the fact that its specific location was chosen exogenously to local natural endowments in an area lacking significant pre-existing populations which mattered at the time of its construction.

The remainder of the paper is structured as follows: I will start-off with summarising the related literature. Then, I will present a detailed description of the underlying data, the history of the TSR as well as the historical trade and post routes linking Europe and China. Third, I will present the empirical models and the results. Section 5 concludes.

2 Related Literature

With the underlying paper I combine findings from the literature on path dependence and transport infrastructure and contribute to a deeper understanding of the question which one of both determinants of economic activity dominates the other one. Scholars attributable to the former

[1] This implies an about 10 times stronger increase in the population of Eastern Russia as compared to European Russia over the same time period.

investigate whether increasing returns or locational fundamentals determine the spatial equilibria of economies. Locational fundamentals being the sole determinant of where economic activity is concentrated would imply that localised shocks would only have short-run consequences which would be set off in the long run. In a recent article, Miguel and Roland (2011) assess the impact of the US bombings during the Vietnam War on subsequent economic activity and find no long-run effect. Their findings are supported by a closely related contribution by Davis and Weinstein (2002) who examine the effects of WW2 bombings in Japan. Further studies in support are Davis and Weinstein (2008) and Miguel and Roland (2011). On the other hand, there is a considerable range of contributions which find evidence for localised shocks to persistently define spatial equilibria in the presence of increasing returns. In a recent contribution, Rauch and Michaels (2013) show that many French towns are still situated in Roman-Age town locations in spite of the fact that those locations often exhibit locational properties inferior to alternatives. Using the construction and the later demise of colonial railroads in Africa, Jedwab and Moradi (2016) show that the associated regional reduction in transport costs had a significant effect on the spatial equilibria during the operation of the railroads. This effect also persisted after their demise. Analysing the spatial equilibrium of the US, Bleakley and Lin (2012) find that modern US cities are often still located close to waterfalls which in the past either caused the foundation of portage cities or provided electricity. Despite the fact that those locational advantages are now obsolete, those cities persisted. Further studies in support of the existence multiple spatial equilibria are among others: Bosker et al. (2007, 2008), Redding et al. (2011), Bleakley and Lin (2012), Rauch and Michaels (2013) and Ahlfeldt et al. (2015).

While both local endowments as well as returns to scale have been found to have significant effects on local development, there also exists a growing literature which puts these into contrast with large scale infrastructure projects. Fogel (1964) pioneered this field of research in being one of the first to apply the social savings methodology to the transport infrastructure field by assessing the impact of railroads on local economic dynamics in the US. Hurd (1983) applied a similar framework to India (Donaldson (2018)). The empirical literature regarding the issue was pioneered by Aschauer (1989). The author was one the first to estimate the relationship between aggregate productivity and stock and flow government spending. Among other things, his findings suggest decisive explanatory power of what he calls core infrastructure projects like highways or mass transit. Duflo and Pande (2007) show that the construction of dams in India increases the agricultural output of downstream districts[2]. In a recent contribution, Donaldson (2018) shows that the construction of the colonial railroad during the British Raj decreased trade costs and increased interregional as well as international trade in India.

In the identification of the causal effects of transport infrastructure on economic development, researchers face a serious problem in the light of the findings which have been laid out previously:

[2] Further contributions assessing the effects of infrastructure projects in different context are: Jensen (2007), Michaels (2008), Dinkelman (2011) and Duranton et al. (2014).

the fact that oftentimes locations which are characterised by either strong localised natural advantages or the presence of increasing returns often exhibit better access to transport networks relative to their counterparts (Duranton et al. (2014)). This opens the issue of endogeneity in econometric models. I contribute to the understanding of the effects of local infrastructure investments by introducing a novel instrumental variable strategy and thereby solving the problem of endogeneity for the underlying empirical model. This allows me to clearly identify the causal impact of the construction of the TSR on local economic development in Russia.

3 Data and Historical Background

Due to the unavailability of data on economic activity in Eastern Russia, I employ data on nocturnal lights emission as provided by the US Airforce's Defence Meteorological Satellite Program (DMSP). Previous studies have established this data as a valid source of information about economic activity (see among others: Michalopoulos and Papaioannou (2013), Storeygard (2016) and Sala-i Martin and Pinkovskiy (2010)). I combine this data with information on the contemporary rail network of Russia as well as with historical data on the tea trade and post routes of Tsarist Russia prior to 1830. My analysis is focused on the part of modern Russia which lies east of the Ural mountains[3]. The reason for that is the fact that this part of modern Russia was - while being largely controlled by Russia - home to less than 300,000 inhabitants until the early 20th century. The data is pre-processed in the following steps:

Lights Data and Local Economic Activity: As there is hardly any reliable information on the location and population of settlements except for the major cities in the Russian East, I follow Jedwab and Moradi (2016) and construct a new data set based on a 0.1x0.1 decimal degree cell grid which covers the entire Russian East. I chose to focus on the area east of the Ural mountains[4] since this area remained largely untouched until the construction of the TSR. Second, as I am interested in the causal effect of the TSR, I further drop all grid cells outside a 500km buffer around the contemporary TSR network and the historical TSR main line, respectively. This leaves me with a data set consisting of about 133,000 grid cells in the contemporary TSR case and roughly 45,000 grid cells in the TSR main line scenario, respectively. As a measure of local economic activity, I extract the mean level nocturnal lights emission per grid cell[5]. As a measure for the local spatial organisation I use the standard deviation in illumination intensity between the 0.1 decimal degree cells which are coded from 0 (no lights) to 63 (sensor satiation) within a 11x11km grid cell.

[3] More specifically, I focus on the area east of the 60.5 longitude line.
[4] This is approximated by dropping all grid cells west of 60.5 degrees longitude.
[5] Please refer to appendix A.1 for a detailed description of the underlying data.

Contemporary and Historical Russian Railroad Network: The - to my knowledge - most comprehensive and accurate as well as freely obtainable shapefile of the Russian Railroad network is provided by diva-gis.org. This shapefile is used in order to compute the shortest distance from each grid cell centroid to the network using ESRI's NEAR tool in arcpy in the full network scenario. Figure I illustrates the underlying algorithm[6]. In order to identify the historical mainline of the TSR, I use the stops summarised in the TSR Wikipedia article[7]. Figures II and III illustrate the respective railroad networks.

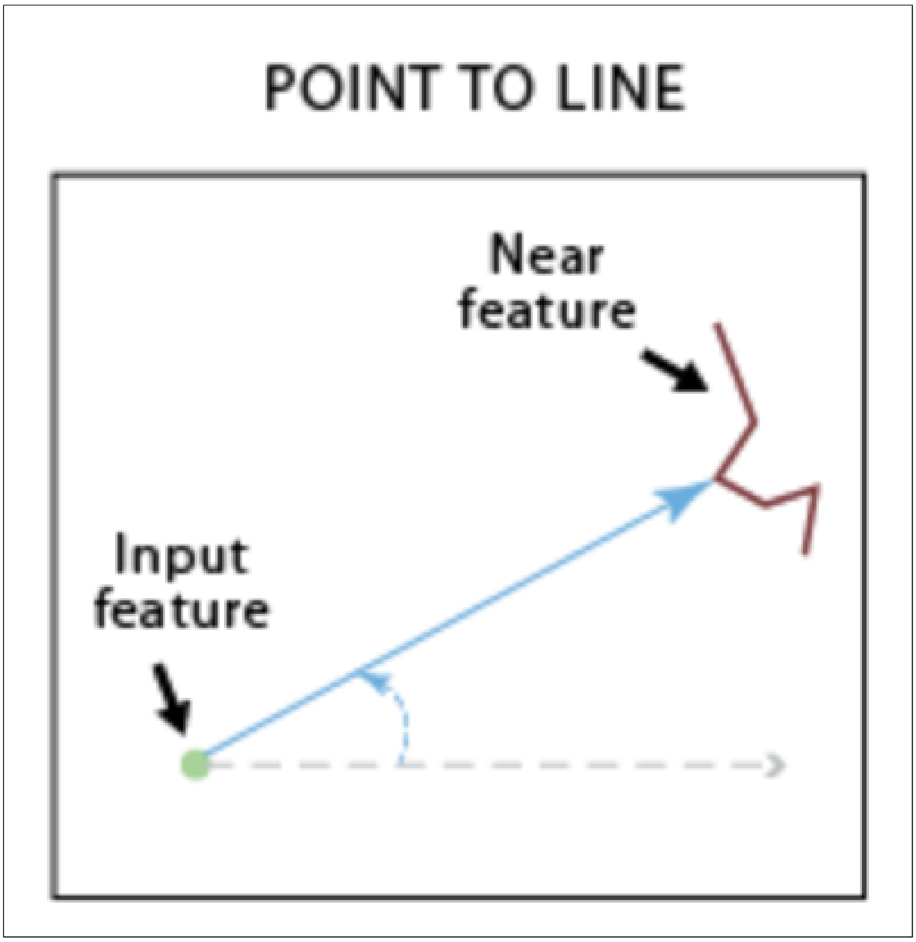

Figure I. Illustration of the NEAR Algorithm by ESRI.

Local (Natural) Advantages: The empirical model laid out in section 4 encompasses several controls for initial local advantages which may act as potential confounding factors. As with the light intensity figures, these data have been extracted for the aforementioned grid cells. These are:

[6] For a detailed description, visit the ArcGIS help page.
[7] The article can be accessed by following this link.

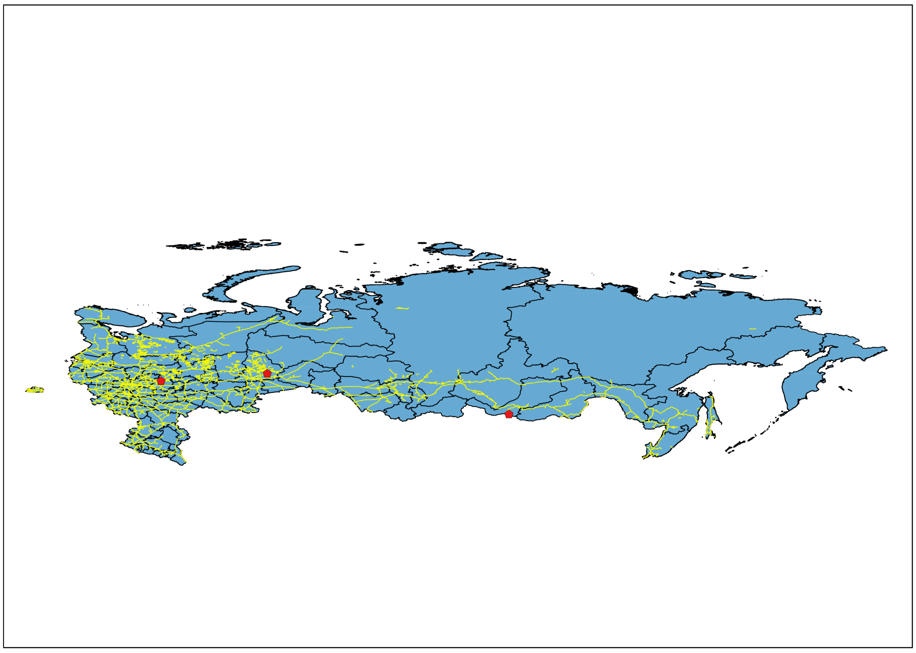

Figure II. The contemporary TSR network.

Caloric potential: the mean local average caloric potential according to the Calorics Suitability Index[8] (CSI) introduced by Galor and Özak (2016) in order to control for local advantages in crop production. This data has been established as valid measure for long run locational advantages by a large number of studies attributable to a school spearheaded by Galor (2011)[9]. As with the night lights data, it is provided in form of raster file out of which I extracted the mean optimal caloric potential per grid cell.

Flare Distance: The minimum distance to the main areas of crude oil and natural gas extraction[10] in order to control for advantages in resource exploitation.

Precipitation: While the average level of precipitation per year usually positively affects the local conditions for crop production on the northern hemisphere, the standard deviation might have adverse effects on both crop production as well settlement suitability due to flooding. The data is extracted from rasters of annual data averaged over 1900-2008[11]. Hence, both measures are included as controls.

Road density: Since local transport infrastructure other than access to the TSR might distort the measurement of the causal effect on economic activity, I include the contemporary

[8] The CSI data can be obtained from Ozak's GitHub Page.

[9] Prominent examples are: Michalopoulos and Papaioannou (2013), Michalopoulos and Papaioannou (2014) and Alesina et al. (2013).

[10] There is no reliable data on the exact location of crude oil and gas extraction sites in Russia available. So I use shapefiles provided by DMSP which encompass the area of gas flares to approximate their position. The shapefiles were obtained from the DMSP data repository.

[11] The data is provided here.

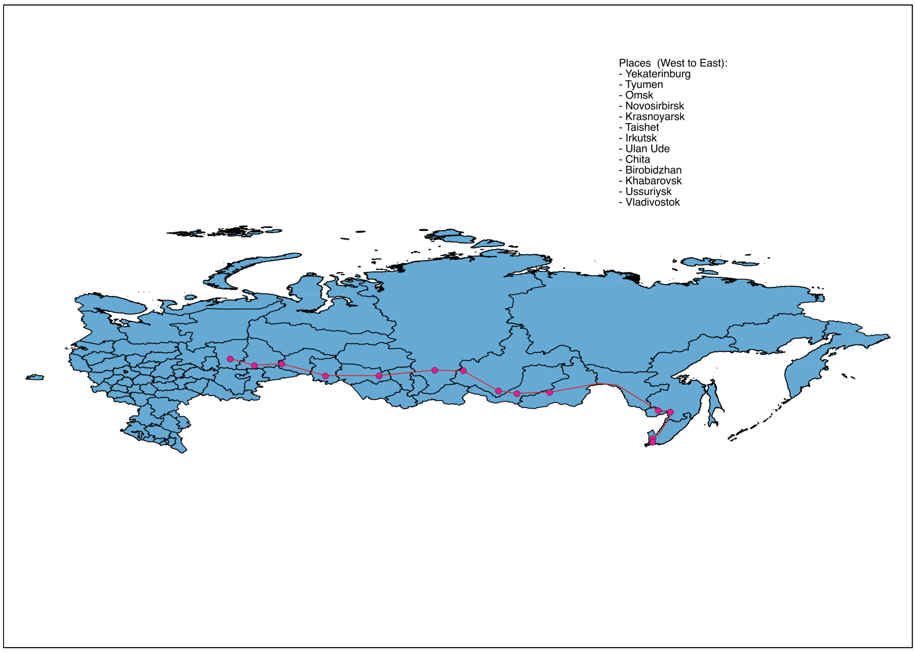

Figure III. The historical TSR mainline.

road density per grid cell as a control variable for local transport costs. This measure is derived from a shapefile encompassing the contemporary road network in Russia[12].

Population density: Aiming to isolate the effect of TSR proximity on economic activity and agglomeration, I control for the population density per grid cell. This measure is derived from the Gridded Population of the World raster data[13].

The Russian Tea Road Tea is one if not the main historical commonality between Russia, China and Central Asia. One of the historically most important trade ports in general and the most important one for the trade between Russia and China instead of being located at the sea shore - as one would intuitively assume taking the more prominent British tea trade as a reference - was located at the Sino-Russian border. The town of Kyakhta was defined as the exclusive border market by the Kyakhta treaty of 1727. Until the outbreak of the Opium War in 1840, Kyakhta and the seaport of Canton were the most prominent Chinese foreign markets (Lee, 2014). In this period, tea was distinguished into two different varieties: Overland Tea which was the one transported via the Russian Tea Route and Canton Tea which was transported by ship from Canton to Europe via the Indian Ocean. The Russian consumers considered the Overland quality to be far superior compared to the sea-borne Canton Tea. This had two reasons. On the one hand, Canton Tea was exposed to hot climate throughout the travel across the Indian Ocean which was said to cause the tea develop an aroma much different to the

[12] The data is provided by DIVA-GIS. I create a density raster based on the shapefile using the Calculate-Density-tool in arcpy. Then I extract the mean road density per grid cell from this raster.

[13] The data is provided by NASA's Socioeconomic Data and Applications Center.

Overland quality. The transport route via the cold and dry deserts of the Russian Tea Route on the other hand was held to be enhancing the taste of the Overland alternative. Lee (2014) states:

> "The unpleasant taste from firing (*bei*) is removed (from tea) by transit through the cold dry climate of Mongolia and Siberia, and at the same time tea, which is nightly unloaded from the camel's back and placed upon the snow-covered steppes, is found to acquire, from light moisture it then absorbs, a delicacy of flavour obtainable in no other way, and it brings, in consequence, a much more lucrative price in the markets of Russia."

Accordingly, there was a strong preference for Overland Tea in the Russian markets where tea was usually consumed without adding milk. While the tea exported to Russia was produced in different provinces throughout China, its entirety was transported to Kyakhta by Chinese traders from where it then was transported to European Russia by camels and oxen cars operated by Russian merchants (Lee (2014)).

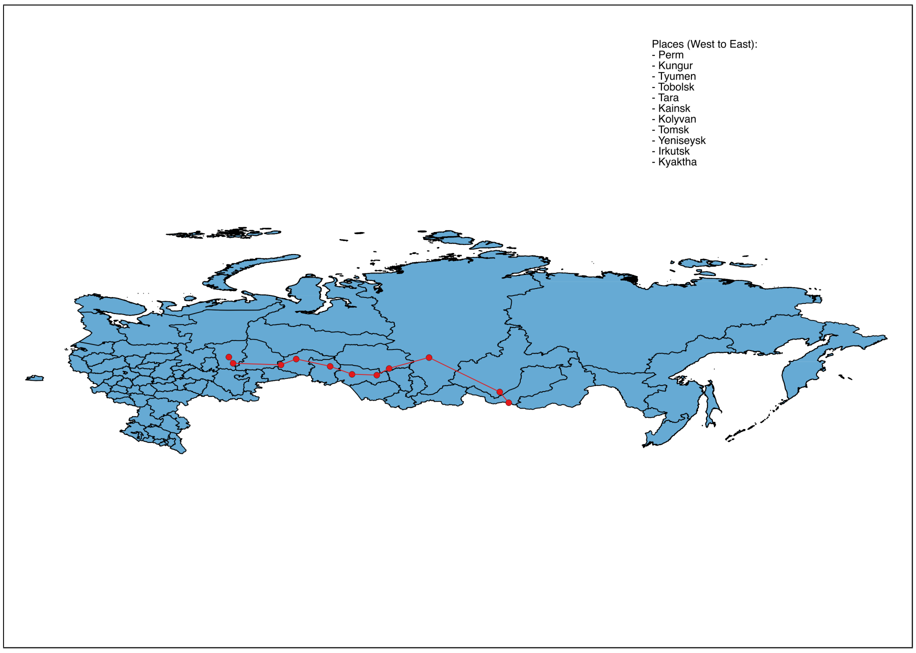

Figure IV. The Tea Route in 19th Century Russia. (Own Depiction based on Avery (2003))

A second function the Russian Tea Route had was serving as a postal road which should ensure communication across the Russian sphere of influence in the east. It extended the Tea Route further east to Vladivostok. The Post Road travelled by Wenyon (1896) is depicted in figure IV[14]. Wenyon who travelled the Russian Post Route in 1893, makes a telling reference to his homeland England in the preface of his book:

[14] The basis for this simplified digitised version is the hand-drawn map by the author which can be found as figure A.2 in appendix A.

“The old post-roads of England have been superseded by the railway, and the same fate will soon befall the great post-road of Siberia.”

The author describes the Siberian Route as a cordon of post-horse stations which were sixteen to twenty miles apart. They were installed by the Russian government for military purposes. Besides those stations he names only a few very scarcely populated places as sources of supply for his travel. He emphasises the notion of the route not being a properly built-up road but at best being a dirt track only defined by the post-horse stations. The author further emphasises his notion of the Route being by far not suitable for the transport of heavy goods or people.

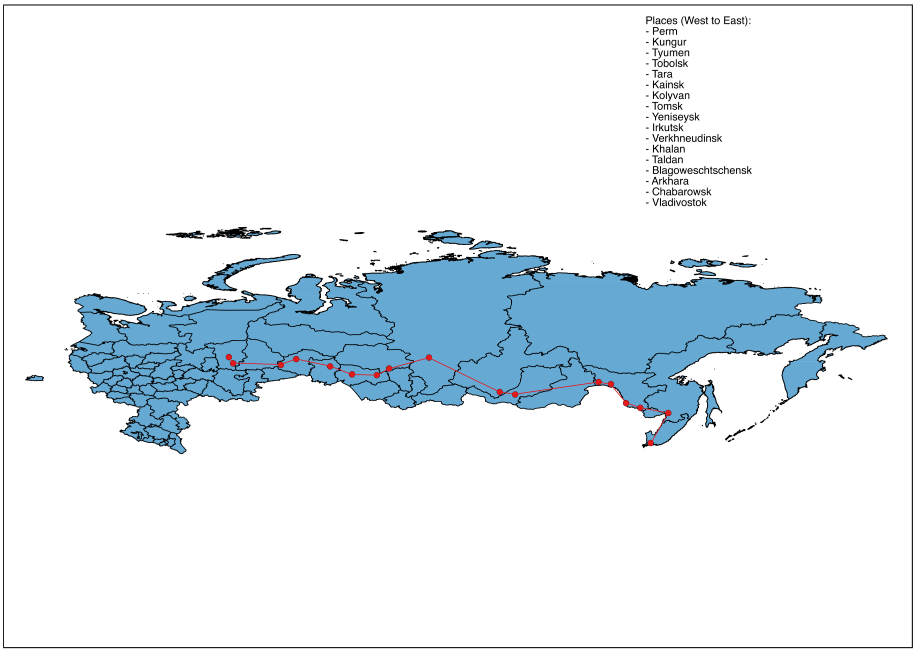

Figure V. The Post Route in 19th Century Russia. (Own Depiction based on Wenyon (1896))

The historical facts which have been laid out in this paragraph strongly support the appropriateness of using the Siberian Route in an instrumental variable approach in the context of my strategy to assess the causal long-run impact of the Russian Railroad on economic activity in Eastern Russia.

4 Econometric Specifications, Identification Strategies and Results

In this section, I will first lay out the basic empirical model used to give a first impression of the positive correlation between TSR proximity and economic activity using simple OLS. Second, I follow Jedwab and Moradi (2016) and use a spatial discontinuity framework to show that the effect found using OLS decrease in the distance to the TSR network. In a third and final step, I propose a novel instrumental variable approach in order to show the causal effect of TSR proximity on local economic development and agglomeration.

4.1 Baseline OLS Regression

The baseline regression model is defined in equations (1) through (4):

$$ln(Lights_i) = \beta ln(DistRail_i) + \chi X_i + \epsilon_i \quad (1)$$

$$ln(Lights_i) = \beta ln(DistMain_i) + \chi X_i + \epsilon_i \quad (2)$$

$$ln(Agglo_i) = \beta ln(DistRail_i) + \chi X_i + \epsilon_i \quad (3)$$

$$ln(Agglo_i) = \beta ln(DistMain_i) + \chi X_i + \epsilon_i \quad (4)$$

where $Lights_i$ is the mean nocturnal lights emission per location or grid cell and agglomeration which is measured by the standard deviation in illumination within each grid cell, respectively. $DistRail_i$ is the variable of interest. It is measured as the geodesic distance between the closest segment of the TSR and the centroid of the location or grid cell, respectively. X_i is a vector of control variables for both local natural advantages as well as current factors which might distort the identification of the local impact of the vicinity to a TSR segment. $DistMain_i$ is the analogue variable of interest in the scenario which focusses on the impact of the distance to the historical TSR mainline. ϵ_i is the error term. Tables I and II give an overview of the employed variables as well as their summary statistics. All variables are in logs with a small number added (0.001) to prevent grid cells with zero lights emission from dropping out of the sample[15].

Table I. Summary Statistics (TSR Scenario)

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	Mean	SD	Min	Max
DistRail	139,871	178.513	146.850	0.000	500.000
DistTea	139,871	1,089.965	881.771	0.004	3,432.420
DistPost	139,871	713.107	605.914	0.001	2,458.482
DistFlare	139,871	543.522	440.308	0.000	1,933.627
Lights	134,878	0.385	2.703	0.000	62.000
CalPot	137,930	1,378.965	1,921.020	0.000	10,827.360
PopDens	139,325	3.833	46.288	0.014	3,121.014
Precip	137,930	446.120	124.803	126.500	1,317.000
PrecipSD	137,930	0.397	1.126	0.000	27.290
DensRoa	139,695	1.962	2.076	0.000	10.282

15 This procedure has - among others - been used in Michalopoulos and Papaioannou (2013) and is widely accepted in the night lights literature.

Table II. Summary Statistics (Mainline Scenario)

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
DistMain	46,140	236.053	143.524	0.001	499.991
DistTea	46,140	277.914	215.171	0.004	1,021.221
DistPost	46,140	207.084	136.814	0.001	672.323
DistFlare	46,140	475.086	279.277	0.000	1,174.234
Lights	45,283	0.744	3.513	0.000	61.618
CalPot	45,773	2,853.748	1,832.917	0.000	8,701.440
PopDens	45,806	8.755	74.513	0.022	3,121.014
Precip	45,773	453.260	103.088	238.000	1,029.000
PrecipSD	45,773	0.565	1.575	0.000	27.290
DensRoa	46,114	3.523	2.473	0.000	10.098

Table III. OLS Results (TSR Scenario)

VARIABLES	(1) Lights	(2) Lights	(3) Lights	(4) Lights	(5) Lights
DistRail	-0.543***	-0.540***	-0.524***	-0.477***	-0.471***
	(0.000)	(0.000)	(0.001)	(0.006)	(0.006)
CalPot		0.039***	0.003***	0.007***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
PopDens			0.231***	0.249***	0.245***
			(0.005)	(0.005)	(0.005)
Precip				-0.086***	-0.095***
				(0.012)	(0.012)
PrecipSD					0.001
					(0.002)
RoadDens					0.026***
					(0.001)
Observations	134,878	133,167	132,843	132,843	132,843
R-squared	0.917	0.919	0.921	0.921	0.921

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equation (1). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the contemporary TSR network.

Tables III through VI report the results from the estimation of equations (1) through (4) using OLS with alternating the dependent variable from mean nocturnal lights emission to the nocturnal lights agglomeration. In both cases, the coefficient while decreasing in magnitude is smaller than zero as expected. The negative conditional correlation between the level of economic activity and the distance to the TSR exhibits only slight changes in magnitude as a response to consecutively including additional controls. The change in magnitude of the coefficient of interest when comparing columns (1) through (6) of table III is below 15 percent. While varying in absolute values, all control variables exhibit the expected negative sign. Comparing the magnitude of the respective coefficients, it is obvious that the negative conditional correlation between the level of nocturnal lights emission and remoteness is the most pronounced. This strong persistence is not observable for the second variant using agglomeration as dependent variable. The coefficient on TSR remoteness decreases by roughly 85 percent comparing (1) through (6).

This change is mainly triggered by the inclusion of the precipitation variable. While it does not break the correlation between agglomeration and TSR access, it strongly diminishes its absolute value. Further it appears that there exists a strong negative relationship between the level of precipitation and agglomeration per grid cell.

Table IV. OLS Results (TSR Scenario)

VARIABLES	(1) Agglo	(2) Agglo	(3) Agglo	(4) Agglo	(5) Agglo
DistRail	-0.525***	-0.523***	-0.504***	-0.154***	-0.141***
	(0.000)	(0.000)	(0.001)	(0.005)	(0.005)
CalPot		0.026***	-0.019***	0.004***	-0.002*
		(0.001)	(0.001)	(0.001)	(0.001)
PopDens			0.292***	0.422***	0.417***
			(0.006)	(0.006)	(0.006)
Precip				-0.651***	-0.662***
				(0.010)	(0.010)
PrecipSD					0.011***
					(0.002)
RoadDens					0.057***
					(0.001)
Observations	139,325	137,606	137,606	137,606	137,606
R-squared	0.892	0.895	0.899	0.903	0.903

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equation (3). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the contemporary TSR network.

Table V. OLS Results (Mainline Scenario)

VARIABLES	(1) Lights	(2) Lights	(3) Lights	(4) Lights	(5) Lights
DistMain	-0.456***	-0.504***	-0.490***	-0.382***	-0.375***
	(0.001)	(0.001)	(0.001)	(0.017)	(0.017)
CalPot		0.106***	0.004***	0.007***	0.005***
		(0.001)	(0.002)	(0.002)	(0.001)
PopDens			0.811***	0.846***	0.836***
			(0.010)	(0.011)	(0.012)
Precip				-0.221***	-0.227***
				(0.034)	(0.034)
PrecipSD					0.012***
					(0.004)
RoadDens					0.026***
					(0.007)
Observations	45,283	44,944	44,636	44,636	44,636
R-squared	0.794	0.803	0.830	0.830	0.831

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equation (2). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

As previously shown, the OLS results suggest a strong negative conditional correlation between the distance to the closest railroad segment and mean nocturnal lights emission as well as

Table VI. OLS Results (Mainline Scenario)

VARIABLES	(1) Agglo	(2) Agglo	(3) Agglo	(4) Agglo	(5) Agglo
DistMain	-0.432***	-0.451***	-0.437***	-0.060***	-0.072***
	(0.001)	(0.001)	(0.001)	(0.017)	(0.017)
CalPot		0.038***	-0.069***	-0.058***	-0.054***
		(0.002)	(0.002)	(0.002)	(0.002)
PopDens			0.841***	0.960***	0.981***
			(0.013)	(0.014)	(0.015)
Precip				-0.769***	-0.742***
				(0.034)	(0.035)
PrecipSD					0.004
					(0.004)
RoadDens					-0.056***
					(0.010)
Observations	45,806	45,465	45,465	45,465	45,465
R-squared	0.758	0.763	0.795	0.798	0.799

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The table reports OLS estimates of equation (4). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

their spatial concentration. Tables V and VI show the results which I receive from conducting the same OLS regression as above but changing the variable of interest to the distance to the TSR mainline. Comparing the coefficients on the variables of interest between the network scenario and the mainline scenario, we observe that changing the specification towards picking up the long run impact of the change in transport costs, they do not considerably change in magnitude if we focus on average night light emission per grid cell. This can be taken as evidence for a stable long run correlation between low transport costs and aggregate economic activity per unit of observation. Looking at the two last columns of table VI, we observe that while diminishing in magnitude, the conditional correlation between vicinity to the TSR mainline and agglomeration of economic activity is negative in the mainline scenario but much less pronounced.

4.2 Further Assessment of the Spatial Equilibrium

After having established a negative conditional correlation between TSR remoteness and economic activity as well as its spatial organisation, I follow Jedwab and Moradi (2016) and estimate equations (5) through (8) in order to assess in how far the installation of the TSR established a specific spatial equilibrium. More specifically, I will lay out the diminishing effect of TSR as one moves further and further away from it.

$$ln(Lights_i) = \beta RailDu_i + \chi X_i + \epsilon_i \tag{5}$$

$$ln(Agglo_i) = \beta RailDu_i + \chi X_i + \epsilon_i \tag{6}$$

$$ln(Lights_i) = \beta MainDu_i + \chi X_i + \epsilon_i \tag{7}$$

$$ln(Agglo_i) = \beta MainDu_i + \chi X_i + \epsilon_i \tag{8}$$

while the dependent variables and the included controls are the same as in the baseline OLS specifications, $RailDu_i$ and $MainDu_i$ are cell dummies which capture the vicinity to the TSR and the TSR mainline, respectively. Those dummies are equal to one if the respective cell lies within 0-10, 10-20, 20-30 or 30-40km distance to the TSR or the TSR mainline, respectively.

Table VII. OLS Results (Vicinity Dummies)

VARIABLES	(1) Lights	(2) Agglo	(3) Lights	(4) Agglo
0-10km Dummy (TSR)	2.5055***	6.9259***		
	(0.0688)	(0.7506)		
10-20km Dummy (TSR)	0.3875***	-0.3789		
	(0.0378)	(0.2426)		
20-30km Dummy (TSR)	0.1394***	-1.1176***		
	(0.0301)	(0.1176)		
30-40km Dummy (TSR)	0.0728***	-1.0330***		
	(0.0247)	(0.0868)		
CalPot	0.0000***	0.0001	0.0001***	0.0002*
	(0.0000)	(0.0001)	(0.0000)	(0.0001)
PopDens	0.0295***	0.4894***	0.0300***	0.5021***
	(0.0012)	(0.0402)	(0.0013)	(0.0448)
Precip	0.0002***	-0.0015***	-0.0006***	-0.0032***
	(0.0000)	(0.0001)	(0.0001)	(0.0008)
PrecipSD	-0.0020	0.0936	0.0175**	0.1148
	(0.0050)	(0.0615)	(0.0070)	(0.0939)
DistFlare	-0.0001***	0.0003***	0.0001*	0.0011***
	(0.0000)	(0.0000)	(0.0000)	(0.0004)
0-10km Dummy (Main)			2.3608***	13.6747***
			(0.2043)	(3.6104)
10-20km Dummy (Main)			1.5268***	6.9187***
			(0.0624)	(1.1122)
20-30km Dummy (Main)			-0.0579	-1.4132
			(0.0558)	(0.8698)
30-40km Dummy (Main)			-0.2619***	-2.0304***
			(0.0419)	(0.5446)
Observations	132,843	137,606	44,636	45,465
R-squared	0.3734	0.5350	0.5291	0.5462

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equations (5) through (8). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR network or the historical TSR mainline, respectively.

Table VII presents the results obtained from estimating equations (5) through (8). Columns (1) and (3) indicate a strong positive and significant effect on the level economic activity of a cell being located relatively close to the TSR or TSR mainline, respectively. This effect diminishes strongly moving towards cells located further away from the TSR. Same holds true for the effect on spatial agglomeration of economic activity within cells as illustrated in columns (2) and (4).

The results presented in this section further substantiate my findings from section 4.1. Still, these results cannot be seen as proof of a causal impact of the closeness to the railroad on economic activity and its spatial organisation within a grid cell as there remains a possibility for the existence of endogeneity. Hence, I propose a novel instrumental variable specification in order to verify the existence of such a relationship. As laid out in section 3, the Transsiberian Routes existed long before the construction of the TSR mainline. And even though, there clearly would have been feasible alternative routes for the TSR since prior to it starting its services the area which is now fairly densely populated was basically empty of people there has been no attempt to relocate it. While the TSR (mainline) was clearly constructed in order to colonise the Russian East, the Transsiberian Routes' locations were not based on existing populations or local natural advantages. This makes them ideal candidates for an IV approach which will be presented in section 4.3.

4.3 Instrumental Variable Approach

Equations (9) through (12) are equivalent to equations (1) through (4) with regards to the included control variables. The specification is estimated using a Two Step Least Squares model where the log of the distance to the nearest Transsiberian Route segment is used as an instrument for the distance to closest TSR (mainline) segment.

$$ln(Lights_i) = \beta(ln(DistRail_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{9}$$

$$ln(Lights_i) = \beta(ln(DistRail_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{10}$$

$$ln(Lights_i) = \beta(ln(DistMain_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{11}$$

$$ln(Lights_i) = \beta ln(DistMain_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{12}$$

Table VIII shows the results for estimating the impact on the mean nocturnal lights emission. The first stage results listed in Panel A show a positive and significant correlation between the logs of the respective distances which while diminishing in magnitude remains positive. This relationship is significant at the 1%-level across all columns. The Kleinbergen-Papp and the Anderson-Rubin-Wald F statistics both indicate that the proposed IV model is not suffering from weak instrument issues[16]. Comparing columns (1) through (7) in panel B, we observe while almost tripling in absolute value, the coefficient on the instrumented variable remains negative

[16] The critical values of the test provided by Stock and Yogo (2005) are 16.38, 8.96, 6.66, and 5.53 for a 10%, 15%, 20%, and 25% bias of the obtained estimator, respectively. Accordingly, the null hypothesis of the underlying estimators being biased due to weak instrumentation are rejected in all cases.

and significant.

Table VIII. IV Results (TSR Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistTea	0.448***	0.445***	0.160***	0.047***	0.052***	0.052***	0.031***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.003)
			Panel B: Second Stage Results				
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistRail	-1.067***	-1.076***	-1.512***	-2.490***	-2.076***	-2.069***	-3.027***
	(0.013)	(0.014)	(0.048)	(0.205)	(0.167)	(0.166)	(0.344)
DistFlare		0.017***	0.016***	-0.013**	-0.031***	-0.031***	-0.008
		(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.005)
CalPot			-0.056***	-0.072***	-0.041***	-0.040***	-0.051***
			(0.005)	(0.008)	(0.006)	(0.006)	(0.009)
PopDens				-0.499***	-0.328***	-0.325***	-0.616***
				(0.077)	(0.062)	(0.062)	(0.118)
Precip					-1.167***	-1.171***	-1.717***
					(0.057)	(0.057)	(0.146)
PrecipSD						0.008***	0.012***
						(0.002)	(0.003)
RoadDens							-0.206***
							(0.030)
Observations	134,878	134,878	133,167	132,843	132,843	132,843	132,843
Kleibergen-Papp LM Stat.	13228	12665	1783	183.2	228	227.9	84.56
Kleibergen-Papp F Stat.	20848	20013	1929	186.7	233.9	233.9	85.83
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (9). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the contemporary TSR network.

Panel B of table IX reports the second stage results for mean nocturnal lights emission as dependent variable and using the Post Route as instrument. The coefficients obtained by IV estimation are all smaller than zero and significant at the 1%-level. Comparing columns (1) through (7), one observes that the coefficient's absolute value decreases by roughly 70 percent. The second stage coefficients are much smaller than in the Tea Route case. This difference stems from a different underlying relationship between the instrument and the instrumented variable due to the fact that the Post Route extends much further east as observable in figure V.

While reacting more strongly to the inclusion of additional controls, the results obtained by instrumental variable estimation suggest a sizeable, significantly negative impact of remoteness to the TSR. All control variables exhibit the expected signs. Further their magnitudes do not change noticeably compared to the OLS results. The picture somewhat changes when we turn to the results from estimating equations (11) and (12). Here I focus on the impact of remoteness relative to the historical TSR mainline. Comparing the first stage results from tables X and XI to the ones from the respective TSR scenarios presented earlier, we observe a stronger and more robust relationship between the instrument and the instrumented variable. This is mainly due

Table IX. IV Results (TSR Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.627***	0.628***	0.436***	0.309***	0.305***	0.305***	0.276***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
			Panel B: Second Stage Results				
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistRail	-0.778***	-0.778***	-0.537***	-0.219***	-0.334***	-0.333***	-0.293***
	(0.009)	(0.009)	(0.016)	(0.024)	(0.025)	(0.025)	(0.028)
DistFlare		0.003	0.005	0.030***	0.013***	0.012***	0.006
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
CalPot			0.040***	0.015***	0.019***	0.019***	0.018***
			(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
PopDens				0.351***	0.317***	0.318***	0.323***
				(0.010)	(0.011)	(0.011)	(0.011)
Precip					-0.673***	-0.679***	-0.628***
					(0.020)	(0.020)	(0.022)
PrecipSD						0.008***	0.007***
						(0.002)	(0.002)
RoadDens							0.030***
							(0.003)
Observations	134,878	134,878	133,167	132,843	132,843	132,843	132,843
Kleibergen-Papp LM Stat.	25945	25897	9577	5814	5253	5252	4216
Kleibergen-Papp F Stat.	41664	41889	12938	7059	6314	6313	5173
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (10). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

to the fact that the historical mainline is closer to the historical routes which is also mirrored in the fact that the differences in the second stage results are by far not as pronounced as in the TSR scenario. Contrasting the insights from both scenarios, we can summarise that there exists a causal negative impact of TSR (mainline) remoteness on local economic activity in Eastern Russia.

Table X. IV Results (Mainline Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistTea	0.456***	0.464***	0.426***	0.381***	0.396***	0.396***	0.382***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
			Panel B: Second Stage Results				
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistMain	-1.039***	-1.043***	-0.697***	-0.400***	-0.398***	-0.397***	-0.384***
	(0.029)	(0.029)	(0.034)	(0.036)	(0.035)	(0.035)	(0.037)
DistFlare		0.019***	0.096***	0.018***	0.019***	0.018***	0.017***
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
CalPot			0.097***	0.007***	0.007***	0.007***	0.005***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.837***	0.838***	0.838***	0.832***
				(0.015)	(0.015)	(0.015)	(0.015)
Precip					0.076	0.062	0.048
					(0.064)	(0.064)	(0.066)
PrecipSD						0.009**	0.009**
						(0.004)	(0.004)
RoadDens							0.022***
							(0.008)
Observations	45,283	45,283	44,944	44,636	44,636	44,636	44,636
Kleibergen-Papp LM Stat.	5882	6062	4954	4683	5110	5106	4794
Kleibergen-Papp F Stat.	6928	7075	5538	5106	5529	5527	4997
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (11). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

$$ln(Agglo_i) = \beta(ln(DistRail_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{13}$$

$$ln(Agglo_i) = \beta(ln(DistRail_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{14}$$

$$ln(Agglo_i) = \beta(ln(DistMain_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{15}$$

$$ln(Agglo_i) = \beta(ln(DistMain_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{16}$$

Turning to the IV results when using spatial agglomeration as dependent variable as I estimate equations (13) through (16), a similar picture arises. As expected, there is no significant difference between the first stage coefficients reported in panel A of table XIII and the ones in table IX. Comparing the coefficients of TSR remoteness from columns (1) through (7), we observe that they are all negative and significant at the 1%-level. Other than the results from estimating the model using the level nocturnal lights emission as dependent variable, the impact of TSR remoteness

Table XI. IV Results (Mainline Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.507***	0.503***	0.473***	0.441***	0.424***	0.424***	0.420***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
			Panel B: Second Stage Results				
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistMain	-0.599***	-0.591***	-0.410***	-0.232***	-0.227***	-0.227***	-0.223***
	(0.028)	(0.029)	(0.031)	(0.030)	(0.032)	(0.032)	(0.032)
DistFlare		0.054***	0.127***	0.029***	0.028***	0.027***	0.024***
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
CalPot			0.112***	0.009***	0.010***	0.009***	0.006***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.890***	0.892***	0.891***	0.874***
				(0.014)	(0.014)	(0.014)	(0.014)
Precip					-0.083	-0.097	-0.105
					(0.064)	(0.064)	(0.064)
PrecipSD						0.010**	0.010**
						(0.004)	(0.004)
RoadDens							0.044***
							(0.008)
Observations	45,283	45,283	44,944	44,636	44,636	44,636	44,636
Kleibergen-Papp LM Stat.	5837	5812	5598	5570	5461	5461	5537
Kleibergen-Papp F Stat.	6058	6019	5599	5410	5167	5166	5145
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (12). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

on lights agglomeration loses only about 30% of its magnitude. The differences between the two instruments in the first stage relationships in the TSR and the mainline scenario differ by similar magnitude as compared to the light emission case. Same holds for the mainline scenario. An interesting insight emerges once comparing the second stage results of the agglomeration specification in both scenarios. Looking at columns 7 in tables XIV and XV, we observe that the coefficient on remoteness shows no statistical significance. While it appears that there is a pronounced causal impact of remoteness on economic activity when focusing on the long run relationship mirrored by the mainline case, agglomeration appears not to be causally impacted by mainline remoteness. Regarding the coefficients obtained for the control variables the picture is highly similar compared to IV results using mean lights emission as dependent variable.

Taking all the above presented results from instrumental variable estimation into account, we can clearly state that there is sizeable positive and highly significant causal effect of being closer to the contemporary TSR network on both the level as well as the concentration of economic activity. While the impact of mainline remoteness on the level of economic activity remains negative and highly significant when focussing on the mainline scenario, this changes when estimate the impact on spatial agglomeration within the respective grid cells. My results suggest that TSR mainline

Table XII. IV Results (TSR Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: First Stage Results			
DistTea	0.445***	0.441***	0.162***	0.046***	0.050***	0.050***	0.033***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
				Panel B: Second Stage Results			
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistRail	-0.949***	-0.996***	-1.464***	-1.824***	-1.633***	-1.630***	-2.187***
	(0.014)	(0.015)	(0.054)	(0.209)	(0.183)	(0.183)	(0.314)
DistFlare		0.069***	0.067***	0.061***	0.057***	0.056***	0.068***
		(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)
CalPot			-0.061***	-0.066***	-0.050***	-0.050***	-0.053***
			(0.005)	(0.008)	(0.006)	(0.006)	(0.007)
PopDens				-0.186**	-0.104	-0.103	-0.274**
				(0.079)	(0.068)	(0.068)	(0.110)
Precip					-0.649***	-0.651***	-0.994***
					(0.063)	(0.063)	(0.134)
PrecipSD						0.003	0.005*
						(0.002)	(0.003)
RoadDens							-0.129***
							(0.025)
Observations	139,325	139,325	137,606	137,606	137,606	137,606	137,606
Kleibergen-Papp LM Stat.	13223	12673	1843	181.9	217.9	217.7	96.38
Kleibergen-Papp F Stat.	20653	20034	1991	185.3	223.3	223.1	97.89
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (13). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR.

remoteness has no significant impact on agglomeration. A potential explanation for this is that it is possible that in the long run centrifugal forces such as congestion economies break the relationship. These results are in line with the theoretical predictions presented in section 1. Further, they support previous empirical research on the topic.

4.4 Validity of the Suggested IV Approach

In order to justify the utilisation of the distance to the historical Transsiberian Route as a valid instrument, the instrument needs to satisfy two conditions according to Cameron and Trivedi (2005): First, it must show a sufficient correlation with the variable of interest X which is to be instrumented. Second, the instrumental variable Z must be exogenous. This means it must not be affected by other variables in the system or in other words: the impact of the instrument on the dependant variable must only be exerted via the instruments impact on the variable which is to be instrumented. The first condition in our case, namely a sufficient (conditional) correlation between the proximity to the TSR (mainline) and the proximity to the Transsiberian Routes is fulfilled as shown by the first stage results as well as the Kleibergen-Paap LM statistics reported in tables VIII through XV. The null of no correlation between the endogenous regressors and

Table XIII. IV Results (TSR Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.622***	0.627***	0.439***	0.311***	0.307***	0.307***	0.280***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
			Panel B: Second Stage Results				
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistRail	-0.816***	-0.800***	-0.723***	-0.407***	-0.460***	-0.459***	-0.446***
	(0.010)	(0.010)	(0.018)	(0.025)	(0.027)	(0.027)	(0.030)
DistFlare		0.064***	0.062***	0.077***	0.074***	0.073***	0.072***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CalPot			0.010***	-0.015***	-0.012***	-0.012***	-0.013***
			(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
PopDens				0.347***	0.333***	0.333***	0.335***
				(0.011)	(0.011)	(0.011)	(0.012)
Precip					-0.322***	-0.325***	-0.307***
					(0.023)	(0.023)	(0.026)
PrecipSD						0.004*	0.003*
						(0.002)	(0.002)
RoadDens							0.010***
							(0.003)
Observations	139,325	139,325	137,606	137,606	137,606	137,606	137,606
Kleibergen-Papp LM Stat.	25612	25814	9788	5922	5365	5363	4408
Kleibergen-Papp F Stat.	41067	41868	13259	7225	6495	6492	5425
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (14). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR.

the excluded instruments is rejected in all cases. Further the test statistics suggest no presence of bias due to weak instruments. This is indicated by the Kleibergen-Paap F statistics. As with the Kleibergen-Paap LM statistics they all are decisively larger than the critical values suggested by Stock and Yogo (2005)[17]. The final potential flaw which is testable, is the failure to fulfill the overidentifying restrictions. As every endogenous regressor is instrumented by exactly one instrument, the equation is exactly identified. Therefore the overidentification restriction is fulfilled. This is also mirrored by the Hansen J statistics reported in tables VIII through XV.

The exogeneity condition can not be tested directly. Still, there are key aspects about the underlying setup in this contribution which strongly support the notion of exogeneity. The first aspect is the considerable time span between the implementation of the Transsiberian Route and the construction of the TSR mainline which amounts to about 180 years as well as the historical contexts of the two events. When the Route was implemented, the Russians subsequently wiped out large parts of the small indigenous population which led to a situation in which Eastern Russia was basically unpopulated until the settlement promoted by the TSR took up. Accordingly, it

[17] The critical values of the test provided by Stock and Yogo (2005) are 16.38, 8.96, 6.66, and 5.53 for a 10%, 15%, 20%, and 25% bias of the obtained estimator, respectively. Accordingly, the null hypothesis of the underlying estimators being biased due to weak instrumentation are rejected in all cases.

Table XIV. IV Results (Mainline Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistTea	0.458***	0.457***	0.420***	0.380***	0.395***	0.395***	0.381***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
			Panel B: Second Stage Results				
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistMain	-0.468***	-0.463***	-0.311***	0.017	0.022	0.022	-0.014
	(0.030)	(0.031)	(0.036)	(0.036)	(0.035)	(0.035)	(0.037)
DistFlare		0.072***	0.089***	0.048***	0.050***	0.050***	0.051***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
CalPot			0.049***	-0.056***	-0.056***	-0.056***	-0.052***
			(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.973***	0.975***	0.975***	0.990***
				(0.017)	(0.017)	(0.017)	(0.017)
Precip					0.150**	0.159**	0.198***
					(0.069)	(0.070)	(0.071)
PrecipSD						-0.006	-0.006
						(0.004)	(0.004)
RoadDens							-0.062***
							(0.011)
Observations	45,806	45,806	45,465	45,465	45,465	45,465	45,465
Kleibergen-Papp LM Stat.	5878	5920	4822	4707	5136	5132	4809
Kleibergen-Papp F Stat.	6962	6951	5403	5124	5548	5546	5015
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (15). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

is highly unlikely that the location of the Route was determined by existing local populations. A second potential concern regarding the exogeneity of the Route's location is the existence of local natural advantages which determined the Route's course. In other words: it could be that the route developed from connecting advantageous places. While theoretically possible, this argument is undermined by two factors. First, as both the Tea Road as well as the Transsiberian Route were mere transport routes for goods and information it appears not logical that local advantages might have played a roll in determining it's location. This is further supported by the results presented in tables XVI and XVII. If the assumption that the course of the routes was not determined by differences in observables, we should not find any once we move further away from the Tea Road or the Post Route, repsectively. This is illustrated by the absence of significant differences in the means of observable controls included in the IV specification as reported in tables XVI and XVII.

Column 4 reports the results from performing t tests on the equality of the means of the respective control variables. I compare the means of the cells which are closer than 10 kilometres to the Tea Road (treated cells) to the ones which are between 10 and 40 kilometres away. As we can see, the null of no differences in the means is rejected in all cases. This can be seen as

Table XV. IV Results (Mainline Scenario)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.510***	0.500***	0.471***	0.443***	0.426***	0.426***	0.421***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
			Panel B: Second Stage Results				
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistMain	-0.262***	-0.216***	-0.164***	0.032	0.024	0.024	0.019
	(0.029)	(0.030)	(0.032)	(0.030)	(0.032)	(0.032)	(0.032)
DistFlare		0.085***	0.099***	0.049***	0.050***	0.050***	0.052***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
CalPot			0.056***	-0.055***	-0.056***	-0.056***	-0.052***
			(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.978***	0.975***	0.975***	0.998***
				(0.016)	(0.016)	(0.016)	(0.016)
Precip					0.148**	0.157**	0.167**
					(0.068)	(0.069)	(0.069)
PrecipSD						-0.006	-0.006
						(0.004)	(0.004)
RoadDens							-0.058***
							(0.011)
Observations	45,806	45,806	45,465	45,465	45,465	45,465	45,465
Kleibergen-Papp LM Stat.	5931	5787	5567	5630	5510	5510	5587
Kleibergen-Papp F Stat.	6117	5946	5529	5448	5196	5195	5174
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (16). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

Table XVI. Summary Statistics (Mean) for Treated and Control Cells (Tea Road)

Group of Cells	0-10 km	10-20 km	10-40 km	t-test (means) 0-10km vs. 10-40km
	(1)	(2)	(3)	(4)
CalPot	3663.119	3637.006	3610.04	0.120
	(30.075)	(29.188)	(16.947)	
DistFla	411.470	411.144	409.958	0.850
	(6.902)	(6.928)	(4.017)	
Precip	454.985	456.071	455.353	0.862
	(1.741)	(1.788)	(1.080)	
PrecipSD	0.554	0.639	0.605	0.349
	(0.044)	(0.049)	(0.028)	
Number of Cells	1,001	1,001	2991	

Note: The table reports the means and standard deviations (in parentheses) for the respective cell groups. Column 4 reports Pr(|T| > |t|) for $H_0! = 0$.

strong support for the assumption that local advantages did not play decisive role in the Tea Road's location. Table XVII reports the means and test statistics for the Transsiberian Route. While differences in the means are slightly different in magnitude as compared to the Tea Road scenario, they still do not exhibit statistically significant differences. Accordingly, this supports the notion of appropriateness of both of the suggested instrumental variable approaches.

Table XVII. Summary Statistics (Mean) for Treated and Control Cells (Transsiberian Route)

Group of Cells	0-10 km	10-20 km	10-40 km	t-test (means) 0-10km vs. 10-40km
	(1)	(2)	(3)	(4)
CalPot	3764.791	3724.435	3666.457	0.092
	(52.873)	(50.541)	(28.730)	
DistFla	580.640	576.656	579.912	0.931
	(7.230)	(7.235)	(4.224)	
Precip	470.179	468.786	468.866	0.649
	(2.510)	(2.504)	(1.441)	
PrecipSD	0.438	0.475	0.464	0.419
	(0.025)	(0.026)	(0.016)	
Number of Cells	1,596	1,591	4,849	

Note: The table reports the means and standard deviations (in parentheses) for the respective cell groups. Column 4 reports Pr(|T| > |t|) for $H_0! = 0$.

Summarising the above explained aspects, all testable conditions of sufficient correlation between the proximity to the TSR (mainline) and the respective instrument variables is fulfilled as shown by the first stage results and the Kleibergen-Paap LM statistics. Potential issues stemming from the failure to fulfill the weak identification as well as the overidentification restrictions are absent as suggested by the Kleibergen-Paap F statistic and the Hansen J statistic. There is ample historical as well as descriptive support for the second, non-testable condition of exogeneity to be fulfilled. Taking all those insights together, there is strong support for the validity of the underlying IV approach.

5 Conclusion

This paper has aimed to contribute to the ongoing scientific debate about the impact of infrastructure projects on local economic development and the local spatial organisation of economic activity. Further, it aimed to empirically disentangle those effects from the ones of existing local natural advantages as well as localised returns to scale by including most controls for both aspects. In doing so, I provide novel cross-section of gridded nightlight emission data, local agro-climatic, resource exploitation in combination with historical data on Tsarist trade and post routes. Due to the (historical) context in which the Transsiberian Railroad was built, I am able to empirically demonstrate a causal negative effect of remoteness to transport infrastructure on local economic activity in Eastern Russia. This effect - while varying in magnitude - is persistent to focussing on either the contemporary TSR network or the historical mainline. In addition to that, I show that this negative causal effect also impacts the local spatial organisation of economic activity when focussing on the contemporary TSR network. This effect vanishes when centering the analysis around the historical mainline which could potentially be explained by centrifugal forces which emerge over time. As a more thorough examination of this question would go beyond the scope of this paper, I relegate it to future research.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6):2127–2189.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, 128(2):469–530.

Aschauer, D. A. (1989). Is Public Expenditure Productive? *Journal of Monetary Economics*, 23(2):177–200.

Avery, M. (2003). *The Tea Road : China and Russia meet across the Steppe*. China Intercontinental Press, Beijing, 1 edition.

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

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2007). Looking for multiple equilibria when geography matters: German city growth and the WWII shock. *Journal of Urban Economics*, 61(1):152–169.

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2008). A century of shocks: The evolution of the German city size distribution 1925–1999. *Regional Science and Urban Economics*, 38(4):330–347.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*.

Davis, D. R. and Weinstein, D. E. (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review*, 92(5):1269–1289.

Davis, D. R. and Weinstein, D. E. (2008). A Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science*, 48(1):29–65.

Dinkelman, T. (2011). The Effects of Rural Electrification on Employment: New Evidence from South Africa.

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

Duflo, E. and Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 122(2):601–646.

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

Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*. Johns Hopkins Press, Baltimore.

Galor, O. (2011). *Unified Growth Theory*. Princeton University Press.

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

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2):994–1028.

Hurd, J. (1983). Railways. In Kumar, D., editor, *Cambridge Economic Histroy of India*, chapter 8. Cambridge University Press, Cambridge, 2 edition.

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

Jensen, R. (2007). The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *The Quarterly Journal of Economics*, 122(3):879–924.

Lee, C. (2014). From Kiachta to Vladivostok: Russian Merchants and the Tea Trade.

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

Michalopoulos, S. and Papaioannou, E. (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, 81(1):113–152.

Michalopoulos, S. and Papaioannou, E. (2014). National institutions and subnational development in Africa. *Quarterly Journal of Economics*, 129:151–213.

Miguel, E. and Roland, G. (2011). The long-run impact of bombing Vietnam. *Journal of Development Economics*, 96(1):1–15.

Rauch, F. and Michaels, G. (2013). Resetting the Urban Network: 117-2012. *Economics Series Working Papers.*

Redding, S. J., Sturm, D. M., and Wolf, N. (2011). History and Industry Location: Evidence from German Airports. *Review of Economics and Statistics*, 93(3):814–831.

Sala-i Martin, X. and Pinkovskiy, M. (2010). African Poverty is Falling...Much Faster than You Think! *NBER Working Papers*, (15775).

Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York.

Storeygard, A. (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies*, 83(3):1263–1295.

Wenyon, C. (1896). *Across Siberia: On the Great Post Road.* Charles H. Kelly, London.

A Appendix

A.1 Using Nightlight Emission as Proxy for Local Economic Activity

One of the main challenges in the underlying project is the lack of reliable sub-national GDP figures for Russia. In their seminal contribution, Henderson et al. (2012) suggest to use the amount of light that can be observed from outer space as proxy for economic activity. While they show that nightlight emission are a viable proxy for economic activity at the national level, the authors further stress that nightlights data is of even greater value in a sub-national setting since it is available at a great geographic fineness of about a 1 square-kilometre resolution. Using this data together with geo-spatial data on for example administrative divisions this data can then be aggregated and be used to construct city or regional-level indices on economic activity. In the following, I will summarise Henderson et al. (2012)'s technical remarks on the data.

The nightlights data is recorded and provided to the public by the United States Air Force Defense Meteorological Satellite Program (DMSP). The program's satellites complete 14 orbits per day since the 1970s. The data is being digitally processed, archived and published since 1992. Originally, the data was being recorded in order to detect clouds and was meant to be used to improve operational accuracy of air strikes and the like. In the process also nocturnal light emissions of human settlements are being recorded as well. The operational pattern of the program ensures that every satellite records every given point on the earth's surface at somewhen between 20:30 and 22:00 o'clock local time. After being transmitted to the program's headquarters, the data is processed by members of the National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center (NGDC). This process entails the removal of unwanted lights emissions like for example forest fires, solar flares or extreme cloud cover with the aim to filter out all natural light emissions or obstructions of of the same. In the end, the cleaned data is aggregated to one composite raster file in order to produce a satellite-year data set which is then made publicly available.

As aforementioned every final product is a raster file which consists of a grid with a 30 arc-second cell size. Such a grid cell approximately covers an area of 0.86 square-kilometres at the equator. The grid extends between 65 degrees south and 75 degrees north latitude. Every grid cell / pixel reports the intensity of nocturnal light on an integer scale from 0 (no lights recorded) to 63 (sensor satiation). While the exclusion of the extreme north and south latitudes clearly means leaving a sizeable portion of the earth's surface uncovered, it is stressed by Henderson et al. (2012) that this area is only inhabited by roughly 10.000 people which is equivalent to 0.0002 percent of the world's population. Accordingly, for the underlying project this means that parts of Russia are left out of the analysis. Still, it is highly unlikely that this significantly distorts the findings. The author's further stress that the recorded night lights reflect all indoor and outdoor use of man-made light. Accordingly, both the use of light in production as well as consumption is recorded and cannot be abstracted. Still, there is a stable relationship between night lights and economic activity which is more than what has been available at this high spatial granularity for Russia. This raw data is then used in order to aggregate the light emission data to less granular grid as described in section 3.

A.2 The Russian Routes Across Siberia

The instrumental variable approaches suggested in this contribution is based on two historical sources. The source used in order to retrace the course and the historical facts regarding the Russian Tea Route is Avery (2003). In her contribution, the author provides a detailed picture about both the known stations of the Tea Route as well its political foundations. Figure A.1 presents the original map which I used in order create a simplified and digitised map. While still being in use, the Tea Route has been extended both its sphere of influence as well as its function after the first operations have been taken up around 1730. As the Russian Tsardom aimed at extending its influence further east, more and more post stations have been installed east of Ulan-Ude. When Wenyon travelled what he then called the Russian Post Route in around 1894, it extended to the most eastern point of the Russian empire Vladivostok. Figure A.2 presents the original scan of the hand-drawn map published in Wenyon (1896). I used this original map in order to create a simplified and digitised version. This version has been used in order to compute the grid centroid distances to the Post Route which I then used in order to instrument the actual distances to the respective rail networks.

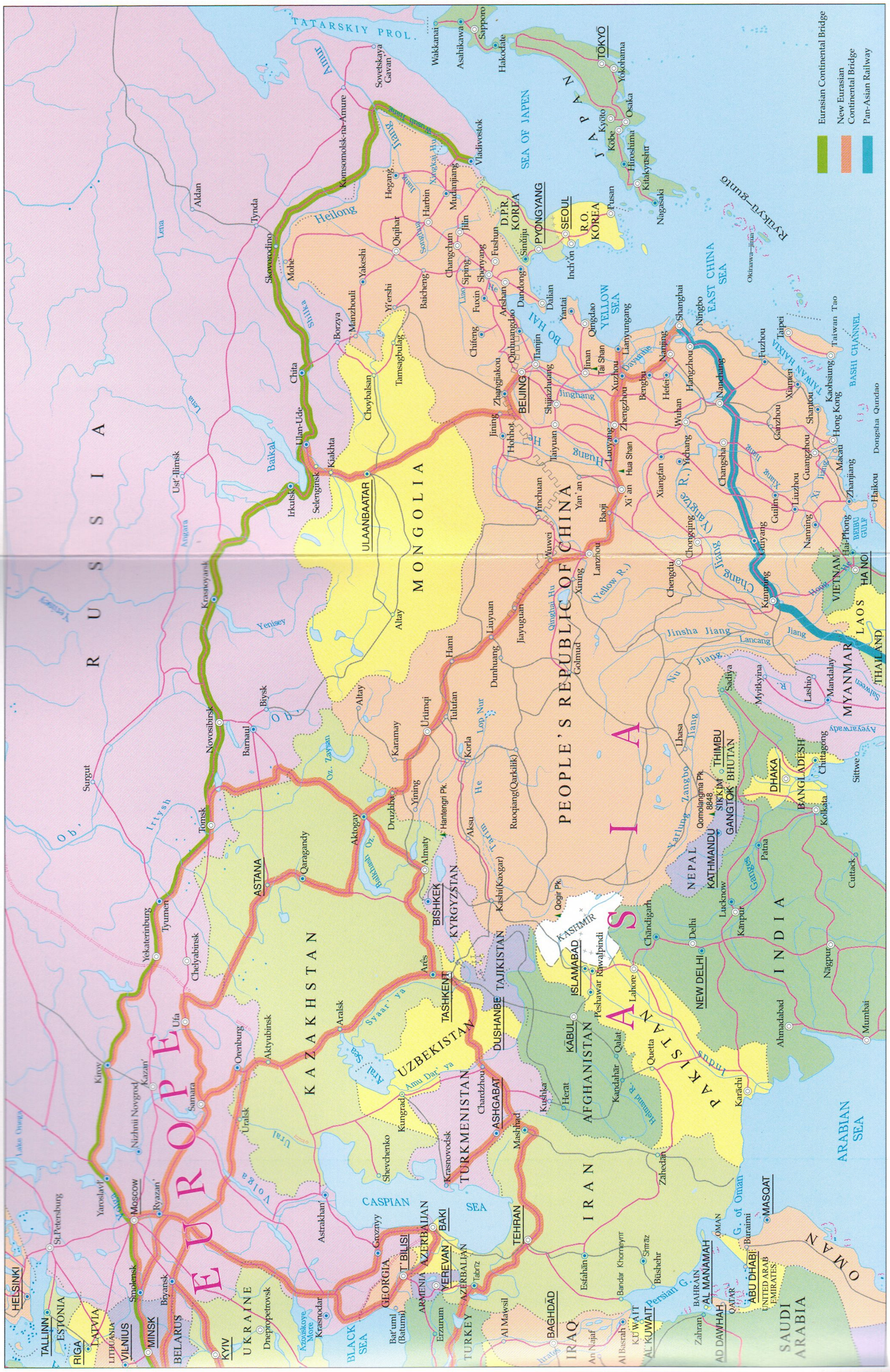

Figure A.1. Tea Road (Avery (2003))

Figure A.2. Post Road (Wenyon (1986))

Hohenheim Discussion Papers in Business, Economics and Social Sciences

This paper series aims to present working results of researchers of all disciplines from the Faculty of Business, Economics and Social Sciences and their cooperation partners since 2015.

Institutes

510	Institute of Financial Management
520	Institute of Economics
530	Institute of Health Care & Public Management
540	Institute of Communication Science
550	Institute of Law and Legal Sciences
560	Institute of Education, Labour and Society
570	Institute of Marketing & Management
580	Institute of Interorganizational Management & Performance

Research Areas (since 2017)

INEPA	"Inequality and Economic Policy Analysis"
TKID	"Transformation der Kommunikation – Integration und Desintegration"
NegoTrans	"Negotiation Research – Transformation, Technology, Media and Costs"
INEF	"Innovation, Entrepreneurship and Finance"

The following table shows recent issues of the series. A complete list of all issues and full texts are available on our homepage: https://wiso.uni-hohenheim.de/papers

No.	Author	Title	Inst
01-2018	Michael D. Howard Johannes Kolb	FOUNDER CEOS AND NEW VENTURE MEDIA COVERAGE	INEF
02-2018	Peter Spahn	UNCONVENTIONAL VIEWS ON INFLATION CONTRAOL: FORWARD GUIDANCE, THE NEO-FISHERIAN APPROACH, AND THE FISCAL THEORY OF THE PRICE LEVEL	520
03-2018	Aderonke Osikominu Gregor Pfeifer	PERCEIVED WAGES AND THE GENDER GAP IN STEM FIELDS	INEPA
04-2018	Theresa Grafeneder-Weissteiner Klaus Prettner Jens Südekum	THREE PILLARS OF URBANIZATION: MIGRATION, AGING, AND GROWTH	INEPA
05-2018	Vadim Kufenko Vincent Geloso Klaus Prettner	DOES SIZE MATTER? IMPLICATIONS OF HOUSEHOLD SIZE FOR ECONOMIC GROWTH AND CONVERGENCE	INEPA
06-2018	Michael Trost	THE WHOLE IS GREATER THAN THE SUM OF ITS PARTS – PRICING PRESSURE INDICES FOR MERGERS OF VERTICALLY INTEGRATED FIRMS	520
07-2018	Karsten Schweikert	TESTING FOR COINTEGRATION WITH TRESHOLD ADJUSTMENT IN THE PRESENCE OF STRUCTURAL BREAKS	520
08-2018	Evanthia Fasoula Karsten Schweikert	PRICE REGULATIONS AND PRICE ADJUSTMENT DYNAMICS: EVIDENCE FROM THE AUSTRIAN RETAIL FUEL MARKET	520

No.	Author	Title	Inst
24-2018	Peng Nie Lanlin Ding Alfonso Sousa-Poza	OBESITY INEQUALITY AND THE CHANGING SHAPE OF THE BODYWEIGHT DISTRIBUTION IN CHINA	INEPA
25-2018	Michael Ahlheim Maike Becker Yeniley Allegue Losada Heike Trastl	WASTED! RESOURCE RECOVERY AND WASTE MANAGEMENT IN CUBA	520
26-2018	Peter Spahn	WAS WAR FALSCH AM MERKANTILISMUS?	520
27-2018	Sophie Therese Schneider	NORTH_SOUTH TRADE AGREEMENTS AND THE QUALITY OF INSTITUTIONS: PANEL DATA EVIDENCE	INEPA
01-2019	Dominik Hartmann Mayra Bezerra Beatrice Lodolo Flávio L. Pinheiro	INTERNATIONAL TRADE, DEVELOPMENT TRAPS, AND THE CORE-PERIPHERY STRUCTURE OF INCOME INEQUALITY	INEPA
02-2019	Sebastian Seiffert	GO EAST: ON THE IMPACT OF THE TRANSIBERIAN RAILWAY ON ECONOMIC DEVELOPMENT IN EASTERN RUSSIA	INEPA

IMPRINT

University of Hohenheim
Dean’s Office of the Faculty of Business, Economics and Social Sciences
Palace Hohenheim 1 B
70593 Stuttgart | Germany
Fon +49 (0)711 459 22488
Fax +49 (0)711 459 22785
wiso@uni-hohenheim.de
wiso.uni-hohenheim.de