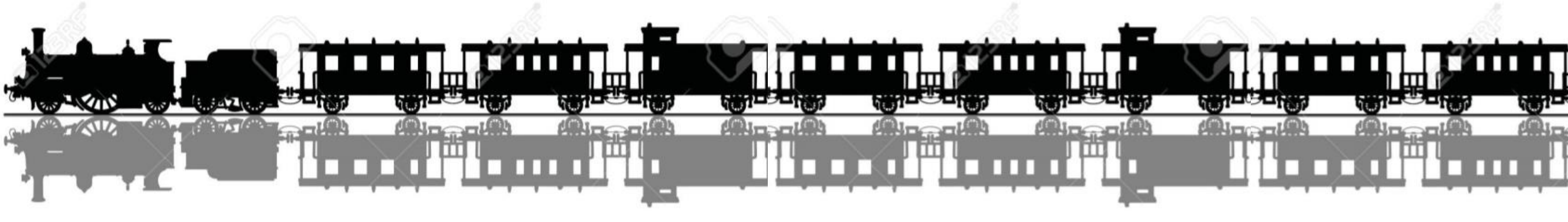
WHERE DO WE LIVE?

An analysis of historical railway development and population distribution in the Netherlands.

Student:	Tim van der Wel
Student number:	2558650
Supervisor:	dr. Hans Koster

VRIJE UNIVERSITEIT AMSTERDAM



Abstract

This thesis aims to find out whether people in the Netherlands have historically moved to areas that were more accessible via rail than others, making those areas significantly denser than other areas. In this thesis, I find that railway accessibility has a spatially heterogeneous and mostly positive effect on historical population density in the Netherlands, albeit small. These results are in line with the theoretical notion of the node-place model, in which the development of a place, and its population density as a result, can be linked to its accessibility.

Contents

Introduction

Since the dawn of the industrial era, rapid changes and technological progress in transport mechanisms have changed the face of the western world. In the past 200 years, wealth has increased with a measure never seen before (Maddison, 2007; Allen, 2009). GDP has increased in every part of the world, and productivity has skyrocketed. These changes, instigated by the industrial revolution, influenced everything, from day-to-day life to the balance of power in the geopolitical world. The aspect of the industrial revolution I focus on in this thesis is the development of rail transport, which began in this era (Maddison, 2007), and its influence on commuting behavior and location choices of people.

In economic literature, the importance of transport is wholly recognized. Transport is vital: not only for the impact it has on the costs of transporting goods, but also for the impact it has on how people travel. Every day, people all over the world transport themselves to a variety of locations, with a variety of goals. Of these transportations, arguably, one of the most impactful is the daily commute. However, while the principle of commuting is not new, as people have always gone from their houses to the location at which they work, the space people traverse in their commute has systematically increased over the past century (Bleijenberg, 2003; Boussauw et al., 2011). In the case of Flanders, Boussauw et al. (2011) found that the average distance individuals commute has increased year after year over the past century. This means either that *'workers have been looking for daily occupations increasingly further away from their home, or – conversely – that they have been moving to a new house further away from their jobs.'* (Boussauw et al, 2011, p.43). Such a finding is in line with the logical thought that if travel costs decrease, then travel consumption increases. This would imply, as a result, a stronger separation between house and job location (Rietveld and Vickerman, 2003; Boussauw et al., 2011).

This thesis aims to find out whether people in the Netherlands have historically moved to areas that were more accessible via rail than others, making those areas significantly denser than other areas. Analyzing population distribution and population density in such a way provides insights into location preferences. However, longitudinal panel data studies researching the effect of transit on population density over time are few and far between. As such, to fill this gap in the literature, this study aims to do precisely such research, using a historical population dataset.

Specifically, this thesis aims to map the historical effects of commuter rail on the allocation of population density from **1831** to **1950**. It is in this period that the industrial revolution in Europe took off, and the period in which commuter rail began to make its appearance in the Netherlands. As such, this thesis is a historical time longitudinal research into the effects of commuter rail on population density and distribution in the Netherlands. The starting year 1831 is chosen as it marked the last year in which the Netherlands did not have an established rail network. Therefore, it is a point of comparison for the following century, in which the Dutch railroads were developed. Similarly, the final year 1950 is chosen to minimize the possible interference from automobile commuting.

Insights into the historical effects of rail on population distributions are not solely academically relevant: in the case of policy regarding transport, such insights may be of interest, too. The historical effect of Dutch rail on population distribution could, for example, influence the decision-making progress governing future locations for railway stations. Additionally, these insights may be used to help designate the suitability of locations for future housing projects.

The structure of this thesis is as follows: first, a theoretical background is given, based on previous literature, historical analyses of transit and contemporary case studies. Second, the methodology of this thesis is expanded upon. Third, the data used is explained, and, fourth, the findings are discussed. Finally, a conclusion is presented.

Theoretical Framework[1]

In this thesis, the idea that individuals prefer to live in accessible locations is based on the node-place model as proposed by Bertolini (2007) and expanded upon by Vale (2015). In this model, land development and transport networks are stated to influence each other through a feedback cycle (Wegener, & Fürst, 2004; Vale, 2015). As it is assumed in this model that a balanced situation between the geographical place and its corresponding accessibility will occur (Vale, 2015), it becomes possible to think of utilizing this theoretical mechanic in spatial economic research. For example, developing an area, and thereby increasing its 'place value', would theoretically automatically lead to a balanced situation in which the 'node value' of that area also increases. By extension, the reverse would also make sense: developing a transport node, and thereby increasing the 'node value' of a location would lead to an increase in the 'place value' of a location (Vale, 2015).

In the case of railway transit and population density, this model implies that an increase in railway accessibility of a location will automatically lead to an increase in population density, and vice versa, until a balanced situation emerges. In other words, I expect that locations typically have a level of railway accessibility and a level of population density, and that these two variables positively affect each other.

[1] This chapter is partly rewritten from an earlier paper of mine, titled*: 'Transit Accessibility and Urban Development: A literature study regarding transit-oriented development and urban development'* for the course 'Research Project, MSC STREEM 2020' at the Vrije Universiteit in Amsterdam.

Also, I propose to theorize population to cluster around transit nodes in a monocentric manner, thereby increasing population density around such nodes.

Including the monocentric city model

In the monocentric city model, a city is regarded as having a central business district where the all the jobs in said city are located, with land value and population density decreasing based on the distance of a given location to said CBD (Alonso, 1964; Wheaton, 1981; Wheaton, 1998). An adapted version of this model could prove a viable way of capturing population distribution and jobs around transit nodes. In this adapted monocentric city model, the central business district would be replaced with the central transport node, assuming jobs in the city to agglomerate around such a central transport node (Wheaton, 2004), and assuming that the transport node stands proxy for the job opportunities in other locations connected to the network.

In other words, if jobs are indeed accessible through the transit network, then it isn't hard to imagine the individual nodes in the network to be proxies for the entirety of available jobs, and as such, as a proxy for the CBD in the classical monocentric city model. In the case of the Randstad area in the Netherlands, Geurs (2006) did in fact find a concentration of jobs around railway stations and in urban areas highly accessible via public transport, implying that such a model is realistic.

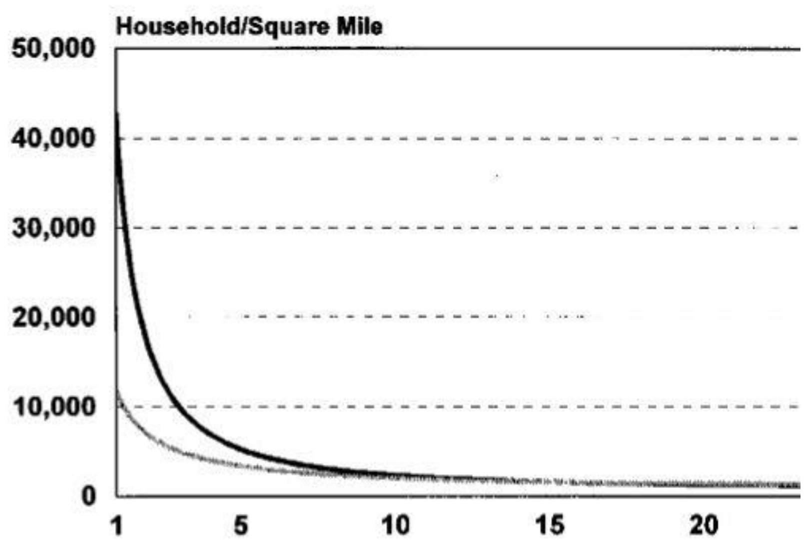


Figure 1. Example of the monocentric city model, after Wheaton (1998)

In figure (1), an example of the monocentric city model has been given. In this specific example, Wheaton (1998) calculated the density of households versus the distance to the CBD in a fictive city. As for the combination of the monocentric city model with the adapted node-place model, I expect population distribution to behave in a similar manner in relation to transit nodes. As such, I expect locations closer to transport nodes to boast a higher level of population and population density, decaying the further locations are from the closest transport node.

If there is any empirical merit in the model I propose, then placing a node somewhere connected to the transit network would be equivalent to placing a CBD in that location, leading to urban development in the area surrounding such a node in the form of increasing population density and absolute population. In terms of historical railway accessibility, this would mean that populations of areas that are well connected grow more rapidly than other areas, and are, therefore, denser.

Previous studies

Railways, railway stations and other modes of public transport and their effects on population density are well-researched. However, the measured effects of public transit on population density are not singular across case studies (Bollinger & Ihlanfeldt, 1997; Levinson, 2008; Bocarejo et al., 2013). In the case of Bogotá, increase in transit accessibility is shown to be a driver for increased population density, with areas closer to transport nodes being denser than before the nodes existed, and denser than comparable, less-accessible areas (Bocarejo et al., 2013). Similarly, in an historical analysis of the relative effects of railway accessibility on the population density in London, Levinson (2008) found that, generally, there is a positive feedback between the two. In this case, Levinson (2008) found that construction of additional transit nodes in led to an increase of population density in the areas around said transit nodes, while an increase in population density led to more transit nodes being constructed. However, this effect was spatially heterogeneous, as the same effect did not occur in the city center of London, where an increase in accessibility led to a rise in commercial development, and subsequently, to depopulation. Levinson (2008) remarks on this matter that one may distinguish between different kinds of density, such as density of jobs, population density and housing density.

In another case study done in Atlanta, U.S.A, (Bollinger & Ihlanfeldt, 1997) found that the effects of relationship between station areas and population where negligible. As Bollinger & Ihlanfeldt (1997) argue, this is because of the high level of car use and the high level of car-based accessibility in the U.S.A, and in Atlanta specifically. In a more global study regarding 621 cities worldwide, Gonzales-Navarro & Turner (2018) find that developing subways in cities do not significantly affect population and population density. In this study Gonzales-Navarro & Turner (2018) look at the general effects of building subways: as such, this specific study does not give insights into the specifics of differences in micro-level areas made more accessible by transit. Still, this finding that city wide density does not significantly change due to the appearance of a subway implies that transit accessibility does not necessarily affect density.

In short, even though the link between transit accessibility and density might seem obvious, literature regarding this subject is divided in its findings. Possible influencing factors include difference in the mediating effect of cultural attitudes towards railway transport, different geographical situations, or other factors.

In the case of commuter rail and its historical effect on population density specifically, findings are more uniform, but literature is scarcer. For example, in their research analyzing the effect of rail on economic growth in the Netherlands between 1840 and 1930, Koopmans et al. (2012) mention that municipalities with lower population densities caught up with bigger and more dense municipalities. Such a result suggests a positive yet unequal effect of historical railway and railway station development in the Netherlands on population density. This notion of a spatially heterogeneous effect of increased accessibility on population density is further corroborated by findings by Baum-Snow et al. (2017). In their 2017 study on contemporary Chinese urban centers, they find that urban railroad and highway accessibility drastically decentralized population from city centers to surrounding prefectures.

Atack et al. (2008, 2009) showed in studies regarding annual population growth in U.S. counties that population growth increased between 1850 and 1860 by 0.41 percentage points if the county had one or more railway lines. In a similar study regarding midwestern U.S. counties in the period between 1840 and 1990, Beeson et al. (2001), too, find a positive effect of railway on population growth. Logically, as increased population growth in a given county is associated with a higher population density in that county, this would mean a historically positive effect of rail on population density. However, in these studies, the positive additional effect of railways found is small compared to general population growth (Atack et al. 2008; Atack et al., 2009; Koopmans et al., 2012).

Based on the consensus of most previous literature regarding historical railways, I expect to find a positive relation between railway stations and density development between the years 1831 and 1950, As such, I hypothesize that historical heightening in population density in the Netherlands is partly caused by railway accessibility. Additionally, based on findings by Koopmans et al. (2012) and Baum-Snow et al. (2017), I further hypothesize such an effect to be spatially heterogeneous, with less urbanized areas being more affected than more urbanized areas.

Design

Data description

In this thesis, I perform a time longitudinal examination of population density in the Netherlands as a function of rail connectivity to jobs. Using historical rail speed approximations, population distribution and network data, I aim to find whether such a relation existed, and how this relation differed between the years 1831 and 1950, with 20-year intervals. In measuring the accessibility to jobs, I use population density as a proxy for jobs in each PC4 area, as population and jobs have often been found to correlate by, among others, Arauzo-Carod (2007) and Steinnes (1982). Using population as a proxy for demand for labor allows for the use of population data to model jobs in different PC4 areas.

The population data of the Netherlands has historically been recorded at the municipality level. However, since municipality borders shift throughout the years, PC4 areas have been chosen as the unit to be observed. PC4 areas are areas associated with the Dutch ZIP code, or postal code. Specifically, PC4 relates to the fourth number in, for example, the postal code 1111. Similarly, PC1 areas relate to the first number in this series. In practice, PC4 areas represent an area typically in between the municipality and neighborhood level in terms of surface size.

Most of the data used is obtained through *the 'Repertorium van Nederlandse gemeenten vanaf 1812'*, translated as *'Directory of Dutch municipalities since 1812'*, compiled by van der Meer & Boonstra (2012). Additionally, the network data used is largely obtained through historical train maps, combined with data from the aforementioned directory.

After gathering the data, I transform the data to make it usable for analysis. First, to get a realistic amount of people that could be reached per year per PC4 area, I utilize geographic information systems to work with spatial population data and to determine population data and density per year per PC4 area in the Netherlands. Second, I generate network rail data for each year. Third, I generate OD matrices that give the accessibility of population via rail for each PC4 in each year. Then, I determine whether a causal relation can be established. As such, after transforming the data, I have a set of observations per PC4 area per year. These observations consist of the population to be reached by foot within 45 minutes and population to be reached by foot and train within 45 minutes. I assume individuals to have a hard limit of 45 minutes that they are willing to spend on a one-way commute based on findings by papers such as Sandow & Westin (2010), and that this preference is consistent throughout the years and places observed. Additionally, I have data on the current-day municipality, as well as the current-day province each PC4 is located in.

Network Data

The reachability of population via train is determined via constructing network data per year researched, based on historical train maps of the Netherlands. The network data consists of maps of the rail system of the Netherlands for the years 1831 – 1950, with 20-year intervals. These were assigned speed values based on the average speed of a train in that year for them to be used in a network analysis using GIS. In addition, a set of maps is provided in the appendix, graphically indicating population density through the years in relation to railway and railway stations (figures 2.1 – 2.7). With these networks and train travel speeds, I determined the amount of people that could be reached per PC4 by rail and foot within 45 minutes. Additionally, I determined the amount of people that could be reached solely on foot in the same time frame, by assuming foot traffic to follow Euclidean paths from each PC4-center to each other PC4-center within 3,75 kilometers.

In mapping network data and accessibility, I assume there to be no additional costs to travelling by rail compared to the alternative of walking. That is, I assume train travel to be free of charge, as is walking. Also, I assume that every station in every year is accessible for passengers (e.g. no purely industrial stations), and that every part of the track is accessible to passenger trains. Finally, I assume that there is no connection to stations and people abroad, and no connection between Dutch stations via foreign soil.

Table (1). Train speeds per year

(1) **Year**	(2) **Average Train Speed**	(3) **Source**
1831	Not relevant (No track in the Netherlands)	-
1850	20 miles per hour -> 32,1869 km/h	White, J. H. (1979)
1870	26 miles per hour -> 41,8429 km/h	White, J. H. (1979)
1890	26 miles per hour -> 41,8429 km/h	White, J. H. (1979)
1910	52,5 km/h	From here on, a linear increase in train speeds is assumed.
1930	65 km/h	A linear increase in train speeds is assumed.
1950	77,5 km/h	A linear increase in train speeds is assumed.

In table (1) above, average real train speeds per year are stated: these could have differed from the average based on terrain, weather, and other factors. Based on these speeds and the networks conceived, the amount of people reachable in 45 minutes by train per PC4 is determined. From 1910 onward, due to a lack of historical sources, I assume a linearly growing net travel speed, increasing with 12,5 km/h per 20 years due to advancing technology. This assumption is based on the notion that by the 1950's, trains with real travel speeds similar to modern trains had made their introduction in Europe (Holland, 2015). Taking this notion and accounting for acceleration time and stops yields the estimate of 77,5km/h real train travel speed for the year 1950. Then, a linearly growing travel speed is assumed between the years 1910 and 1950. This net travel speed, again, includes stops and their frequency in the average speed per year.

Regarding table (1), since the most reliable data regarding real train speeds before the start of the 20th century stems from research into rail and trains in the U.S.A, where such travel speeds differed from those in Europe (White, 1979), these speeds are partly based on the assumption that real travel speeds by rail in Europe were similar to real travel speeds in the USA.

Summary Statistics

In table (2) below, a summary of the data used is represented on the PC4 area level. Additionally, summary statistics split by year (table 5), as well as graphical representations of the railway networks per year (figures 2.1 – 2.7) can be found in the appendix.

Table (2) Summary Statistics[2]

	(1) N	(2) mean	(3) sd	(4) min	(5) max
Population	27345	16351	2629	1	85018
Area of PC4 (km2)	27345	8.453	9.218	0.0228	115.5
Population density per km2	27345	448.8	1415	0.143	30669
Population in walking distance	27345	14800	33573	4	478742
Population reachable by train	27345	125597	752606	0	1.365e+07
Accessibility train and walk combined	27345	138488	760961	4	1.37e+07
Population density per km2 (log)	27345	4.841	1.314	-1.945	10.33
Population in walking distance (log)	27345	8.632	1.246	1.386	13.08
Population reachable by train (log)	27345	3.077	4.819	0	16.43
Accessibility train and walk combined (log)	27345	9.005	1.779	1.386	16.43

In the Netherlands, 4032 PC4 areas exist today, in varying sizes. Not considering PC4 areas which did not exist in all of the years researched reduces this total to 3938. Some PC4 areas not existing in all of

[2] Note that models (1.1) and (1.2) are left out of all the regression tables. This is done as these regressions are uninformative due to an omitted variable bias.

the years researched is mostly due to the province Flevoland not existing until the 1960's (Bakker, 1957). Note that PC4 areas which did exist in some years but not in others are, in fact, considered. Such areas are, for example, PC4's in the Haarlemmermeerpolder, which was not laid dry until the middle of the 1850's (Taverne, 2006). In such cases, the observations in years in which these PC4 areas did not exist have been removed, while observations in other years have been maintained.

In total, 27345 observations per variable were collected, found in table (2). Note that the logarithmic transformations of some variables are given. This is done to ensure a normal distribution of data despite outliers, and to make regression results more interpretable. Histograms of the logarithmic variables by year, as well as scatterplots of 'population density' and 'population reachable by train' and 'population reachable by train and foot' are stated in the appendix (figures 3 and 4).

The minimum value of the variable 'Population' being 1, as found in table (2), is odd. However, this can be explained through the fact that some PC4 areas overlap with both inhabited as well as uninhabited areas through the years. As such, in the case of PC4's consisting of "new land", won from damming in the water, there could be some spillover of registration of population from older, neighboring PC4 areas. A good example is the case of Schokland, a former island in what was known as the Zuiderzee, now part of Flevoland. Due to a spatial spillover of population, the neighboring PC4's of Schokland all have population counts ranging from 1 to 6. However, this data is valid, and as such, it is kept in the data analysis, as these people did, technically, live in these PC4 areas in the years researched.

Model

In determining the effects of railway accessibility on population distribution, I construct the following model:

$$\log(Y_{ti}) = \beta_0 + \beta_1 \log(x_{ti}) + \varepsilon_{ti} \qquad (1.1)$$

Adding in control variable $\log(a_{ti})$ gives:

$$\log(Y_{ti}) = \beta_0 + \beta_1 \log(x_{ti}) + \beta_2 \log(a_{ti}) + \varepsilon_{ti} \qquad (1.2)$$

in which dependent variable Y_{ti} represents population density in PC4 *i* and year *t*, with x_{ti} representing the independent variable of the amount of people accessible in 45 minutes via both train and foot. a_{ti} is a control variable, capturing the amount of people that can be reached on foot in 45 minutes. This accounts for a "natural" accessibility of areas, such as areas which are close to, or are themselves, urban centers with a high level of population and population density.

Regressions (1.1) and (1.2) do pose a problem: they do not consider possible hidden variables which could influence the population density in given PC4 areas in given years. This problem is known as the omitted variable bias. To combat such a possible omitted variable bias, I propose some multiple fixed effects models, in which I include both location and time fixed effects. Based on model (1), I propose three such multiple fixed effects models: one in which I include two-way fixed effects on both province and year, one in which I add PC4 fixed effects, and one in which I add a Year*Province-number interaction variable fixed effect. Including year and place fixed effects account for time and place invariant heterogeneities between PC4 areas and provinces, respectively. Additionally, the inclusion of Year*Province interaction variable fixed effects account for effects which exist per province, but which vary per year. Examples of variables which correlate with population density, and where such fixed effects models account for, are water accessibility and location specific government policies (Coale, 1981; Small & Nicholls, 2003).

First, I estimate model (2.1) in which I fix effects of both the province level as well as on time:

$$\log(Y_{ti}) = \gamma_t + \theta_p + \beta_1 \log(x_{ti}) + \varepsilon_{ti} \tag{2.1}$$

Adding in control variable $\log{(a_{ti})}$ gives

$$\log{(Y_{ti})} = \gamma_t + \theta_p + \beta_1 \log{(x_{ti})} + \beta_2 \log{(a_{ti})} + \varepsilon_{ti} \tag{2.2}$$

where γ_t and θ_i represent time and province fixed effects, respectively. Additionally, Y_{ti} represents the population density, x_{ti} represents the population reachable in 45 minutes via both train and foot and a_{ti} represents population distance in walking distance across the PC4's $i = 1, \dots, n$ for the years $t = 1831, \dots, 1950$.

Second, I construct model (3) in which PC4 fixed effects are added, in addition to year and province fixed effects. This is done to find more detailed location specific fixed effects:

$$\log(Y_{ti}) = \gamma_t + \theta_p + \varphi_i + \beta_1 \log{(x_{ti})} + \beta_2 \log{(a_{ti})} + \varepsilon_{ti} \tag{3}$$

where γ_t and φ_i represent year and PC4 fixed effects, respectively. All other parameters are the same as in equations (2.1) and (2.2).

Third, I construct model (4) in which year fixed effects, province fixed effects, PC4 fixed effects and interaction variable Year*Province fixed effects are added:

$$\log(Y_{ti}) = \gamma_t + \theta_p + \varphi_i + \omega + \beta_1 \log{(x_{ti})} + \beta_2 \log{(a_{ti})} + \varepsilon_{ti} \tag{4}$$

where γ_t, φ_i and ω represent time fixed effects, PC4 fixed effects and Year*Province interaction fixed effects, respectively. All other parameters are the same as in equations (2.1), (2.2) and (3).

In all models 1 through 4, I cluster standard errors on the 2009 municipality level, as the original data as constructed by van der Meer & Boonstra (2012) was gathered at that level. Clustering standard errors at this level is prudent as in this case, as the municipality data can be regarded as a clustered sample (Abadie, Athey, Imbens & Wooldridge, 2017).

Results and Discussion

In this paragraph, the statistical findings and regressions run based on the models discussed are presented in table (3), and a sensitivity analysis is done in table (4). Additionally, two variants of model (2.2) and one of model (3), in which the effects of population accessible in 45 minutes on population density are split by year and by province, can be found in tables (6), (7) and (8) in the appendix.

Table (3) Regression results

Dependent: Population per KM2

	(1) 2FE Model 2.1	(2) 2FE Model 2.2	(3) 2FE Model 3	(4) 2FE Model 4
Population accessible in 45 minutes by train and foot (log)	0.437***	0.0451***	-0.00270	-0.00524
	(0.0324)	(0.0120)	(0.00741)	(0.00853)
Population in walking distance (log)		0.767***	1.030***	1.001***
		(0.0267)	(0.0351)	(0.0337)
Year Fixed Effects	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES
PC4 Fixed Effects	NO	NO	YES	YES
Year*Province Fixed Effects	NO	NO	NO	YES
Observations	27,196	27,196	27,178	27,178
R-squared	0.496	0.652	0.921	0.922

Robust standard errors in parentheses, clustered on 2009 municipality level
*** p<0.01, ** p<0.05, * p<0.1

Table (3) above shows that in model (2.1), the measured effect is 0.437% increase in population density for each 1% increase in population accessible by train and foot in 45 minutes. When adding the control variable 'population in walking distance' in model (2.2), the coefficient changes to a 0.0451% in population density for each 1% increase in population accessible by train and foot in 45 minutes. This means that adding in the control variable ‘population in walking distance’ reduces the measured effect by a factor 10.

Seeing as the control variable ‘population in a 45-minute walk’ is highly correlated with ‘population density’, is significant at the 1% level, and is responsible for a large decrease in in the measured effect of the variable ‘Population accessible in 45 minutes by train and foot (log)’ of population density, I find that this variable should indeed bae included in any further regressions. As can be seen in tables (6) and (7) in the appendix, the positive effects that model (2.2) finds are consistent across most individual years and provinces.

In model (3), I find a 0.00270% decrease in population density for each 1% increase in population accessible by train and foot within 45 minutes. However, these findings are not significant. In model 4, the coefficient states that a 1% increase in population reachable by train and foot in 45 minutes is associated with a 0.00524% decrease in population density, again not significant. When comparing models (3) and (4), I find that adding in Year*Province interaction effects in model (4) does not change the findings much compared to model (3). This implies that either there were relatively few time-variant province-fixed effects observed, or that these effects were not impactful.

What stands out is the fact that I find mostly positive effects, with varying degrees of significance per model used and fixed effects included: when adding PC4 fixed effects in model (3), the effect of ‘population reachable by train and by foot’ changes to an insignificant coefficient. As such, there is strong reason to suspect that there are PC4 fixed effects which highly influence population density. Such PC4 fixed effects could be a level of preexisting urbanization, jobs, or amenities in certain PC4 areas.

When looking at tables (6), (7) and (8) in the appendix, I see that, when splitting model (2.2) per year and per province, most effects hold per province and per year: in almost all instances, the effect found here is positive and significant, albeit small. When splitting model (3) per province, I find that, again, the findings in table (3) hold most of the time: most coefficients found are insignificant.

Extension of the models

Some issues with all regressions presented in table (3) could stem from endogeneity issues. Most importantly, denser and more populous areas, consisting mostly of pre-established cities, could attract infrastructure in the form of rail connections, thereby reversing the effects between population density and accessibility. In order to account for this possible endogeneity issue, I aim to apply the *inconsequential place approach* as described by Chandra & Thompson (2000) and Redding & Turner (2015), which relies on choosing a sample that is *‘inconsequential in the sense that unobservable attributes do not affect the placement of infrastructure’*. (Redding & Turner, 2015, p. 1368). This accounts for the notion that railways were historically constructed to connect more populous cities with each other: in these cases, it is the population and population density which attracted rail connections, and not vice versa. As such, to negate the effect of reverse causality due to rail being laid between cities that already have higher population and population density, I propose to remove larger cities from the estimations, based on their 2019 population levels (CBS, 2020).

First, I remove the largest 10 cities, being Amsterdam, Rotterdam, The Hague, Utrecht, Eindhoven, Tilburg, Groningen, Nijmegen, Enschede and Haarlem from the observed variables (CBS, 2020), thereby negating the interaction between accessibility and population density in these cities. As a result, the effects observed more reliably show the effect of increased rail accessibility on population density. Secondly, I perform a variant of this inconsequential place approach with the largest 20 cities which exist in all years observed. These cities include, in addition to the aforementioned 10 cities, Breda, Arnhem, Zaanstad, Amersfoort, Apeldoorn, 's-Hertogenbosch, Maastricht, Leiden, Dordrecht and Zoetermeer (CBS, 2020). Additionally, I perform a third variant in which the next 10 largest cities are also removed, being Zwolle, Deventer, Delft, Alkmaar, Heerlen, Venlo, Leeuwarden, Hilversum, Hengelo and Amstelveen (CBS, 2020).

Table (4). Sensitivity analysis: Inconsequential place analysis

Dependent: Population per KM2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	2FE Model 2.2 10 cities removed	2FE Model 3 10 cities removed	2FE Model 4 10 cities removed	2FE Model 2.2 20 cities removed	2FE Model 3 20 cities removed	2FE Model 4 20 cities removed	2FE Model 2.2 30 cities removed	2FE Model 3 30 cities removed	2FE Model 4 30 cities removed
Population accessible in 45 minutes by train and foot (log)	0.0461***	0.00857*	0.00861*	0.0420***	0.0109**	0.0109**	0.0428***	0.0108**	0.0111**
	(0.0128)	(0.00480)	(0.00480)	(0.0135)	(0.00481)	(0.00491)	(0.0143)	(0.00506)	(0.00520)
Population in walking distance (log)	0.706***	0.962***	0.950***	0.694***	0.962***	0.953***	0.645***	0.953***	0.943***
	(0.0292)	(0.0318)	(0.0367)	(0.0320)	(0.0325)	(0.0362)	(0.0285)	(0.0349)	(0.0386)
Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
PC4 Fixed Effects	NO	YES	YES	NO	YES	YES	NO	YES	YES
Year*Province Fixed Effects	NO	NO	YES	NO	NO	YES	NO	NO	YES
Observations	24,301	24,284	24,284	22,878	22,861	22,861	21,667	21,650	21,650
R-squared	0.614	0.939	0.939	0.602	0.943	0.943	0.585	0.937	0.937

Robust standard errors in parentheses, clustered on 2009 municipality level
*** p<0.01, ** p<0.05, * p<0.1

The first thing that stands out in table (4) above is that the effects found in model 3 and 4 are positive, but very small across the board. Adding in PC4 fixed effects makes the measured effects much smaller relative to adding in only province and year fixed effects. Removing more cities seems to make these positive effects more significant, up to the 5% level. Again, effects seem to change most when including PC4 fixed effects in all regressions run. Interesting is the finding that, after removing the 20 most populous cities from the analysis, the coefficient remains significant and positive, and does not change much when removing additional cities. This implies that it is mostly these 20 largest cities which react insignificantly to increased accessibility. Also, this implies that there was indeed a measure of reverse causality between population density in bigger cities and rail accessibility.

The shift in sign of the coefficients found when removing larger cities from the regressions implies that PC4 areas in larger cities largely experience no effect on density due to an increase in accessibility, while PC4 areas outside such cities generally experience an increase in population density. This last notion is in line with Koopmans et al. (2012), who mention that municipalities with lower population densities caught up with bigger and more dense municipalities due to increased accessibility in a historical analysis of Dutch rail.

Such a finding is interesting: it suggests that PC4 areas that were not located in bigger cities did experience an increase in population density. This would mean that people historically moved to areas which became more accessible due to the construction of railways, and which were not dense before. In comparison, cities did not experience the same effects, suggesting that the added accessibility did not influence individuals to move to such cities.

Conclusion

The models presented and regressions run in this thesis quantify the historical effect of railway accessibility on population density. Generally, when effects are significant, I find small positive effects of population accessibility via train and foot on population density. However, the regressions run imply a spatially heterogeneous effect: removing bigger cities leads to more positive and more significant effects. Such findings are in line with most previous literature, such as papers Atack et al. (2008, 2009), Koopmans et al. (2012) and Beeson et al. (2001): railway accessibility is generally found to have a small, positive and spatially heterogeneous effect on historical population density in the Netherlands.

These results are partly in line with the theoretical notion of the node-place model, in which the development of a place, and its population density as a result, can be linked to its accessibility. Also, these findings are in line with the hypothesis of this thesis, namely there being a spatially heterogeneous and positive effect of railway accessibility on population density between 1831 and 1950 in the Netherlands.

When looking at the spatial heterogeneity of the effects found, I find that it is generally locations outside of populous urban areas which benefit from an increase of rail accessibility in terms of increasing population density. If this spatial heterogeneity is indeed due to a displacement effect as described by Baum-Snow et al. (2017), then such a finding would imply increasing rail accessibility to lead to suburbanization. As such, such spatial heterogeneity of effects could be associated with a stronger separation between house and job location, an effect previously described in papers by Rietveld & Vickerman (2003) and Boussauw et al. (2011).

The main purpose of the research done in this paper was to provide insights into the historical location preferences of people in relation to railway accessibility in the Netherlands. In that, this thesis succeeded. These insights could be used for, for example, weighing the importance of increasing accessibility and developing logistical projects aimed at doing so. However, the effects found are mostly very small. Compared to other countries, these findings imply that the Netherlands are no exception in how accessibility influences population distribution in the period researched: generally, small and positive effects are found.

I propose that further research into this topic focusses on the development of the effect of rail accessibility on population density after 1950. Such research should consider other forms of transport that rose in significance in this period, such as the car. Also, seeing the development of effects of the train through the years could grant insights into the shifting preferences of people to live in cities or suburbs, among other developments. Additionally, performing similar research in other countries would grant insights in whether the effects found in this paper are location specific.

References

Literary references

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). When should you adjust standard errors for clustering? (No. w24003). National Bureau of Economic Research.

Allen, R. C. (2009). Engels' pause: Technical change, capital accumulation, and inequality in the British industrial revolution. Explorations in Economic History, 46(4), 418-435.

Alonso, W. (1964). Location and land use. Toward a general theory of land rent.

Arauzo-Carod, J. M. (2007). Determinants of population and jobs at a local level. *The Annals of Regional Science*, *41*(1), 87-104.

Atack, J., Haines, M.R., Margo, R.A., 2008. Railroads and the rise of factory evidence for the United States. 1850–70, NBER Working Paper 14410.

Atack, J., Haines, M.R., Margo, R.A., 2009. Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest. 1850–60, NBER Working Paper 14640.

Bakker, D. (1957). Oostelijk Flevoland raakt begroeid. *De Levende Natuur*, *60*(12), 305-310.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. Review of Economics and Statistics, 99(3), 435-448.

Beeson, P.E., DeJong, D.N., Troesken, W., 2001. Population growth in U.S. counties, 1840–1990. Regional Science and Urban Economics 31, 669–699.

Bleijenberg, A. (2003). The driving forces behind transport growth and their implications for policy. Managing the fundamental drivers of transport demand, 37.

Bocarejo, J. P., Portilla, I., & Pérez, M. A. (2013). Impact of Transmilenio on density, land use, and land value in Bogotá. *Research in Transportation Economics*, *40*(1), 78-86.

Bollinger, C. R., & Ihlanfeldt, K. R. (1997). The impact of rapid rail transit on economic development: The case of Atlanta's MARTA. *Journal of Urban Economics*, *42*(2), 179-204.

Boussauw, K., Derudder, B., & Witlox, F. (2011). Measuring spatial separation processes through the minimum commute: the case of Flanders. *European journal of transport and infrastructure research. -Delft*, *11*(1), 42-60.

Chandra, A., & Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, *30*(4), 457-490.

Coale, A. J. (1981). Population trends, population policy, and population studies in China. Population and Development review, 85-97.

Geurs, K. T. (2006). Job accessibility impacts of intensive and multiple land-use scenarios for the Netherlands' Randstad Area. *Journal of Housing and the Built Environment*, *21*(1), 51-67.

Gonzalez-Navarro, M., & Turner, M. A. (2018). Subways and urban growth: Evidence from earth. *Journal of Urban Economics*, 108, 85-106.

Holland, J. (2015). *The Train Book: The Definitive Visual History*. Dorling Kindersley.

Koopmans, C., Rietveld, P., & Huijg, A. (2012). An accessibility approach to railways and municipal population growth, 1840–1930. *Journal of Transport Geography*, *25*, 98-104.

Keijer, M. J. N., and Piet Rietveld. "How do people get to the railway station? The Dutch experience." *Transportation planning and technology* 23.3 (2000): 215-235.

Levinson, D. (2008). Density and dispersion: the co-development of land use and rail in London. *Journal of Economic Geography*, *8*(1), 55-77. `

Maddison, A. (2007). The world economy volume 1: A millennial perspective volume 2: Historical statistics. Academic Foundation.

Rietveld, P. (2000). The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D: Transport and Environment*, *5*(1), 71-75.

Redding, S. J., & Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. In *Handbook of regional and urban economics* (Vol. 5, pp. 1339-1398). Elsevier.

Sandow, E., & Westin, K. (2010). Preferences for commuting in sparsely populated areas: The case of Sweden. *Journal of Transport and Land Use*, *2*(3/4), 87-107.

Steinnes, D. N. (1982). Do people follow jobs or do jobs follow people'? A causality issue in urban economics. *Urban Studies*, *19*(2), 187-192.

Small, C., & Nicholls, R. J. (2003). A global analysis of human settlement in coastal zones. Journal of coastal research, 584-599.

Taverne, E. (2006). Het betwiste landschap van de Haarlemmermeer. *BMGN-Low Countries Historical Review*, *121*(4), 711-726.

Wheaton, W. C. (1998). Land use and density in cities with congestion. *Journal of urban economics*, *43*(2), 258-272.

Wheaton, W. C. (2004). Commuting, congestion, and employment dispersal in cities with mixed land use. *Journal of Urban Economics*, *55*(3), 417-438.

White, J. H. (1979). *A History of the American Locomotive: Its Development, 1830-1880*. Courier Corporation.

Data References

CBS (2020), *Bevolkingsontwikkeling; Bevolking aan het einde van de periode december 2019,* Retrieved on 30-06-2020 from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37230ned/table?dl=318D6

Hollandse IJzeren Spoorweg-Maatschappij (1889), *Net der Hollandsche IJzeren Spoorweg-Maatschappij,* Retrieved on 30-04-2020 from: *http://www.nicospilt.com/index_spoorkaarten.htm*

Kaart der Nederlandsche Spoorwegen en Tramwegen (1934). *Uitgave der Nederlandsche Spoorwegen,* Retrieved on 30-04-2020 from: *http://www.nicospilt.com/index_spoorkaarten.htm*

Spoorwegkaart van Nederland (1910), *Spoorwegkaart van Nederland*, Uitgegeven bij H. ten Brink*, Meppel,* Retrieved on 30-04-2020 from: *http://www.nicospilt.com/index_spoorkaarten.htm*

Spoorkaart 1949 (1949). Retrieved on 30-04-2020 from: *http://www.nicospilt.com/index_spoorkaarten.htm*

Van den Henvell & van Santen (1868), *Lijnen zoals voorkomende in: Van den Henvell & van Santen's Officieele Reisgids, Officieele Reisgids,* Retrieved on 30-04-2020 from: *http://www.nicospilt.com/index_spoorkaarten.htm*

Van der Meer & Boonstra, *Repertorium van Nederlandse gemeenten vanaf 1812,* DANS data guide, Retrieved on 17-06-2020 from:
https://dans.knaw.nl/nl/over/organisatiebeleid/publicaties/DANSrepertoriumnederlandsegemeenten2011.pdf

Appendix

Summary statistics split by year

Table 5: Summary statistics split by year

1831

	N	mean	sd	min	max
Population	3449	591.163	754.969	1	16045
Area in KM2	3486	9.146	9.453	.095	115.523
Population per km2	3449	114.635	357.536	.163	11262.27
Population walk center	3480	3906.792	4692.1	4	39455
Population reachable by foot and train	3480	3906.792	4692.1	4	39455
log Population per km2	3449	4.095	.973	-1.814	9.329
log Population walk center	3480	7.845	.919	1.386	10.583
log Population reachable by foot and train	3480	7.845	.919	1.386	10.583

1850

	N	mean	sd	min	max
Population	3456	687.876	836.002	1	16496
Area in KM2	3486	9.146	9.453	.095	115.523
Population per km2	3456	130.833	375.373	.163	11573.56
Population walk center	3482	4500.48	5154.806	5	48244
Population reachable by foot and train	3482	6508.042	31082.55	5	629654
log Population per km2	3456	4.267	.951	-1.814	9.356
log Population walk center	3482	8.014	.891	1.609	10.784
log Population reachable by foot and train	3482	8.045	.963	1.609	13.353

1870

	N	mean	sd	min	max
Population	3477	803.011	975.376	2	20757
Area in KM2	3489	9.14	9.45	.095	115.523
Population per km2	3477	152.466	409.285	.327	12571.51
Population walk center	3487	5307.392	5939.956	18	60194
Population reachable by foot and train	3487	32057.15	189000	18	2076401
log Population per km2	3477	4.458	.9	-1.118	9.439
log Population walk center	3487	8.199	.854	2.89	11.005
log Population reachable by foot and train	3487	8.413	1.265	2.89	14.546

1890

	N	mean	sd	min	max
Population	3470	944.112	1204.392	1	32129
Area in KM2	3482	9.159	9.453	.095	115.523
Population per km2	3470	183.234	479.481	.143	14066.62
Population walk center	3480	6512.561	8086.573	20	102728
Population reachable by foot and train	3480	70655.87	384000	20	4226930
log Population per km2	3470	4.596	.974	-1.945	9.552
log Population walk center	3480	8.353	.908	2.996	11.54
log Population reachable by foot and train	3480	8.767	1.517	2.996	15.257

1910					
Population	3470	1156.2	1463.983	1	41127
Area in KM2	3482	9.159	9.453	.095	115.523
Population per km2	3470	235.021	517.692	.143	10350.1
Population walk center	3480	8585.614	12194.32	21	152026
Population reachable by foot and train	3480	169000	745000	21	1.01e+07
log Population per km2	3470	4.784	1.032	-1.945	9.245
log Population walk center	3480	8.552	.967	3.045	11.932
log Population reachable by foot and train	3480	9.228	1.811	3.045	16.131
1930					
Population	3482	1522.433	1825.738	1	50306
Area in KM2	3494	9.135	9.446	.095	115.523
Population per km2	3482	319.371	581.244	.286	6320.436
Population walk center	3492	11771.97	16208.09	27	145765
Population reachable by foot and train	3492	326000	1270000	27	1.37e+07
log Population per km2	3482	5.073	1.048	-1.252	8.752
log Population walk center	3492	8.824	1.009	3.296	11.89
log Population reachable by foot and train	3492	9.654	1.981	3.296	16.431
1950					
Population	3497	1994.845	2497.478	2	68383
Area in KM2	3508	9.134	9.444	.095	115.523
Population per km2	3497	415.7	723.661	.286	9153.123
Population walk center	3508	15450.24	21052.52	30	201514
Population reachable by foot and train	3508	195000	1010000	30	8270021
log Population per km2	3497	5.324	1.071	-1.252	9.122
log Population walk center	3508	9.07	1.041	3.401	12.214
log Population reachable by foot and train	3508	9.456	1.665	3.401	15.928

Maps

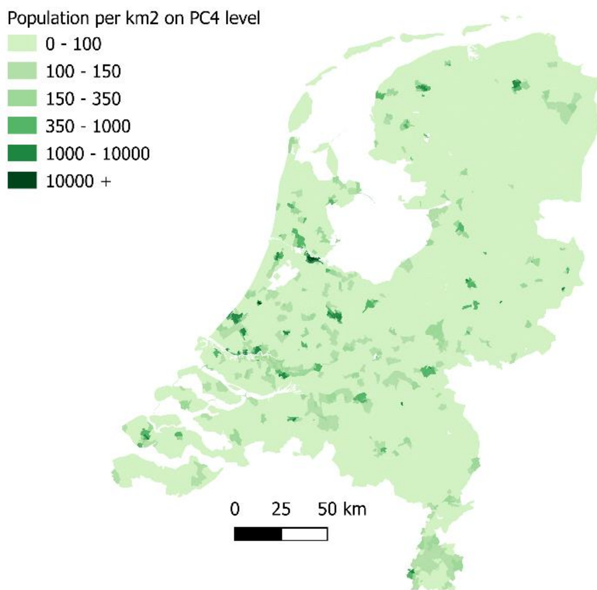


Figure 2.1. Map of population density in 1831

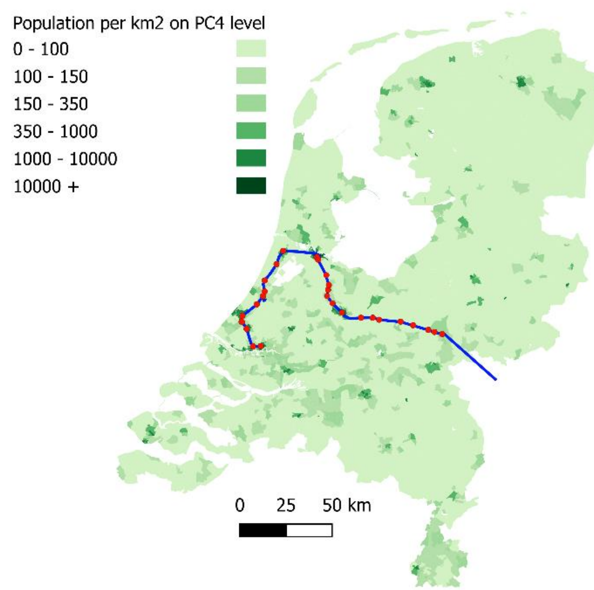


Figure 2.2. Map of population density in 1850

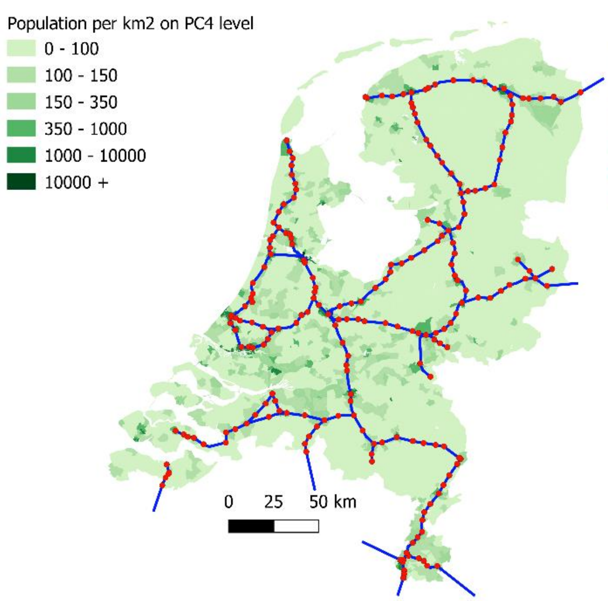


Figure 2.3. Map of population density in 1870

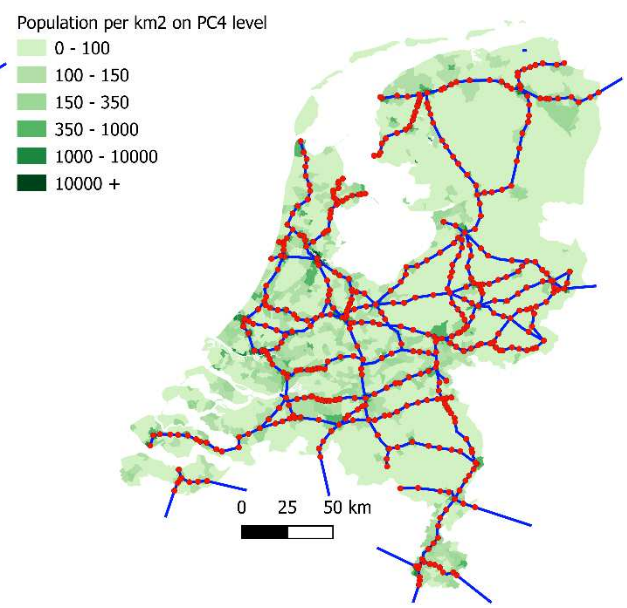


Figure 2.4. Map of population density in 1890

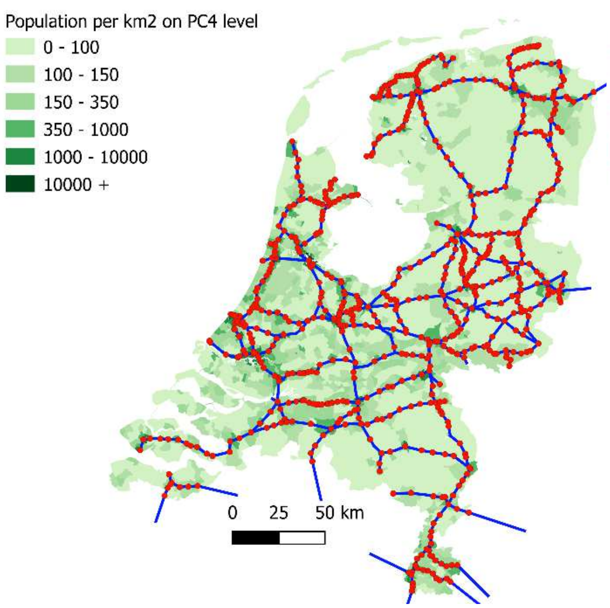


Figure 2.5. Map of population density in 1910

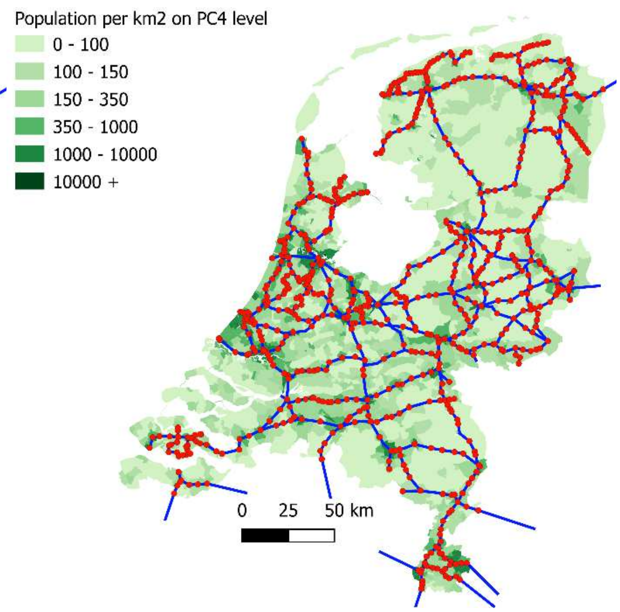


Figure 2.6. Map of population density in 1930

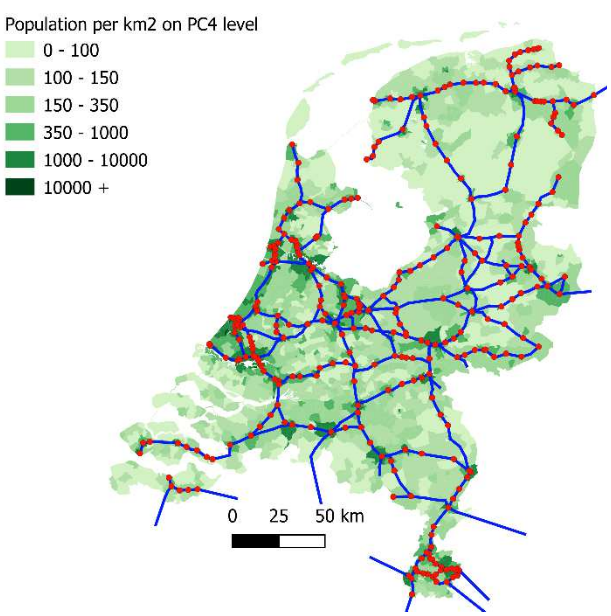


Figure 2.7. Map of population density in 1950

Histograms

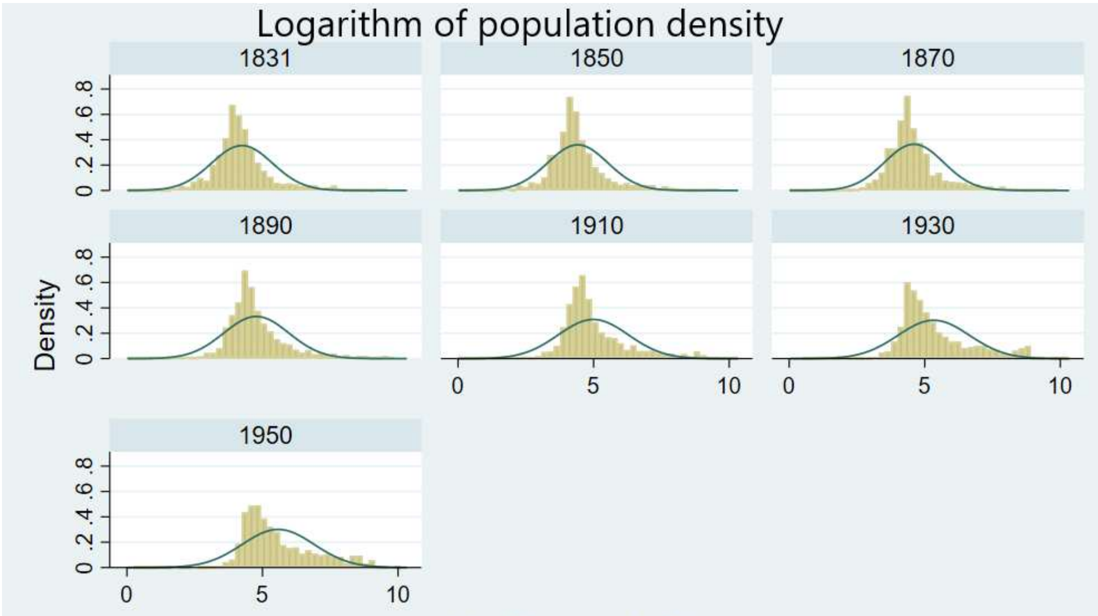


Figure 3.1 - Histograms of log-transformed population density per km2 per year.

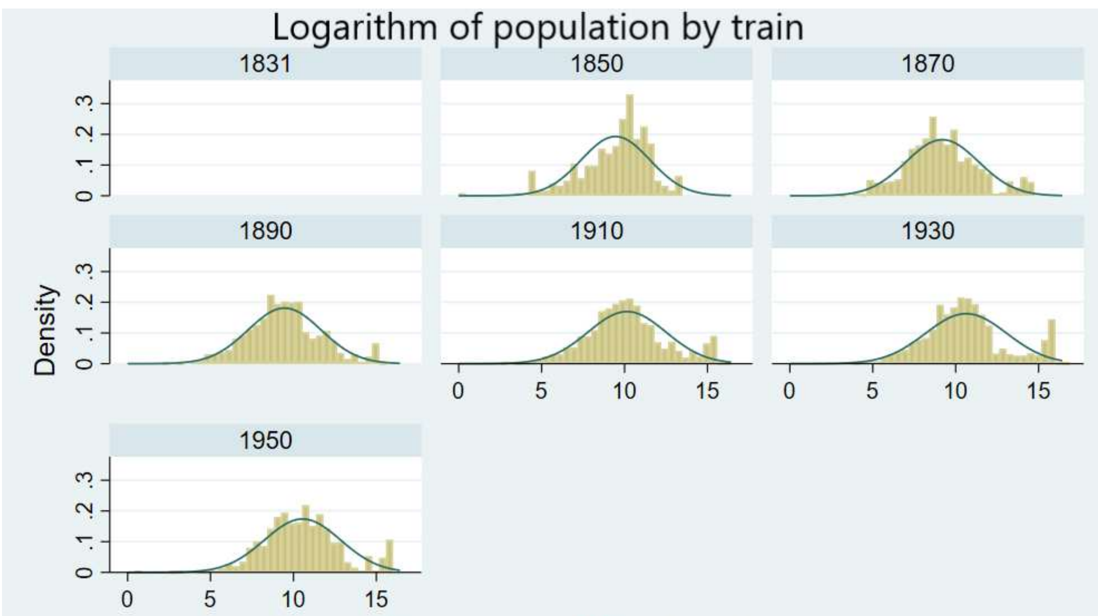


Figure 3.2 - Histograms of log-transformed population reachable by train per year.

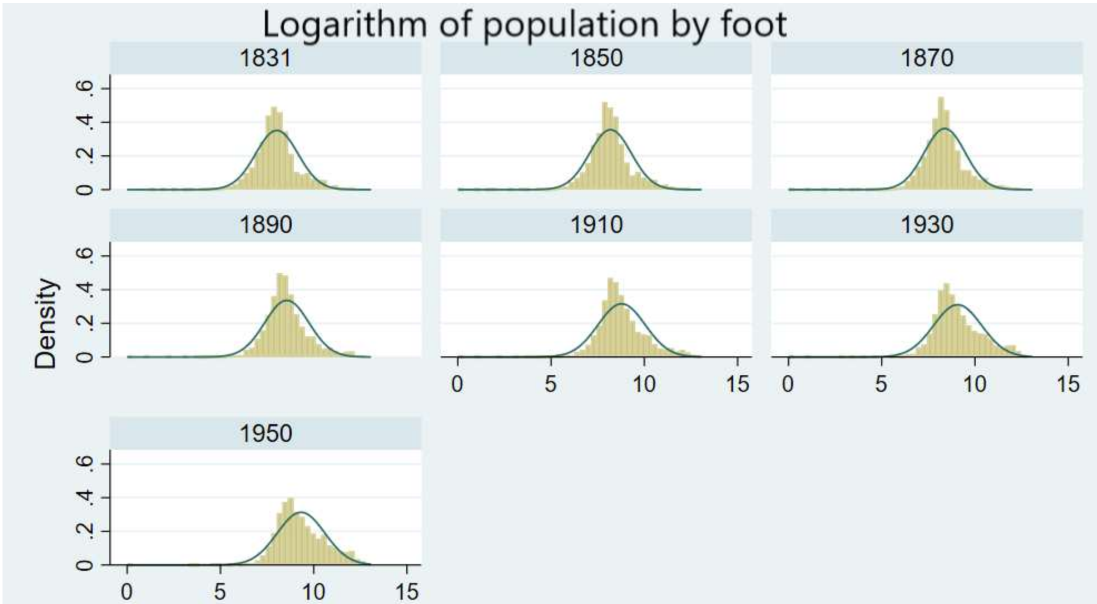


Figure 3.3 - Histograms of log-transformed population reachable by foot per year.

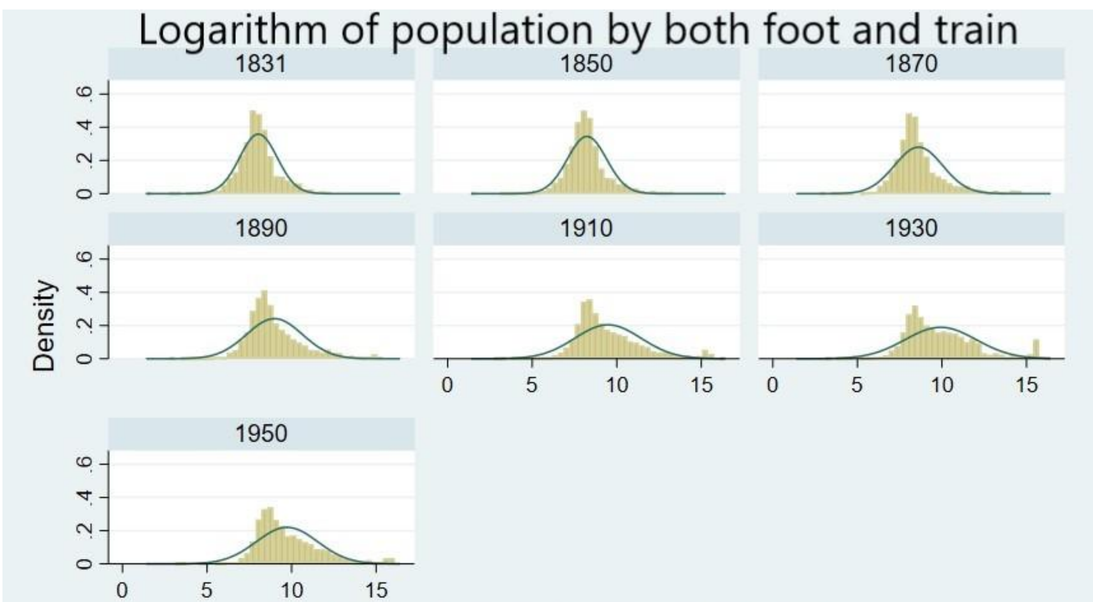


Figure 3.4 - Histograms of log-transformed population reachable by train and foot combined per year.

Scatterplots

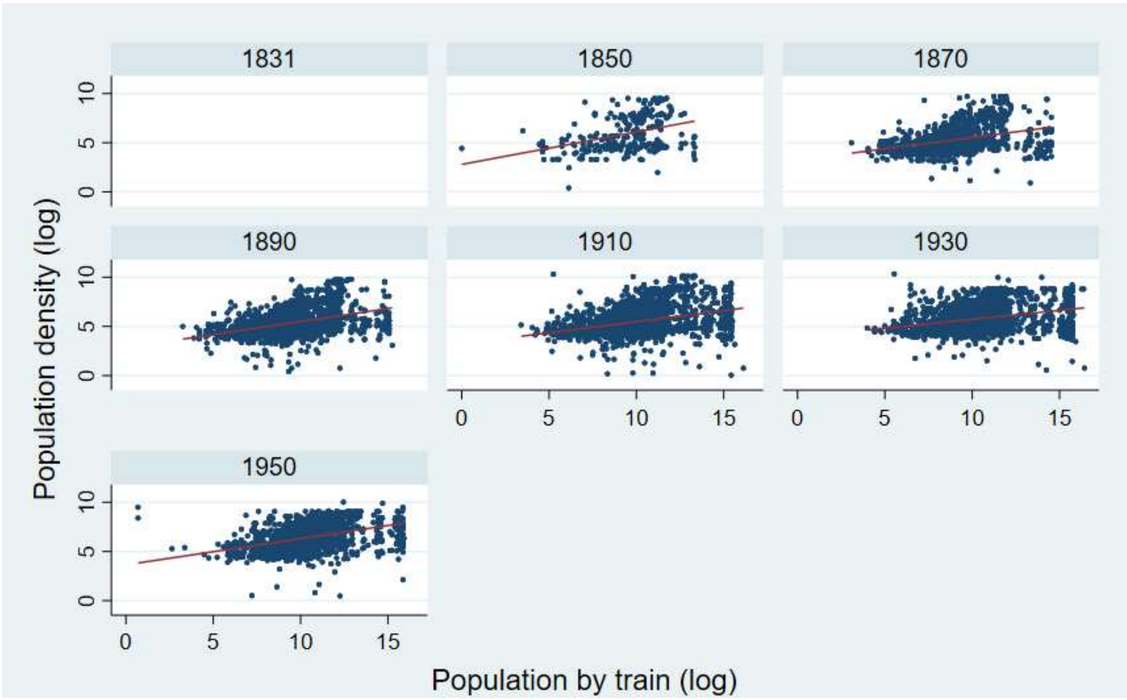


Figure 4.1. Scatterplots by year

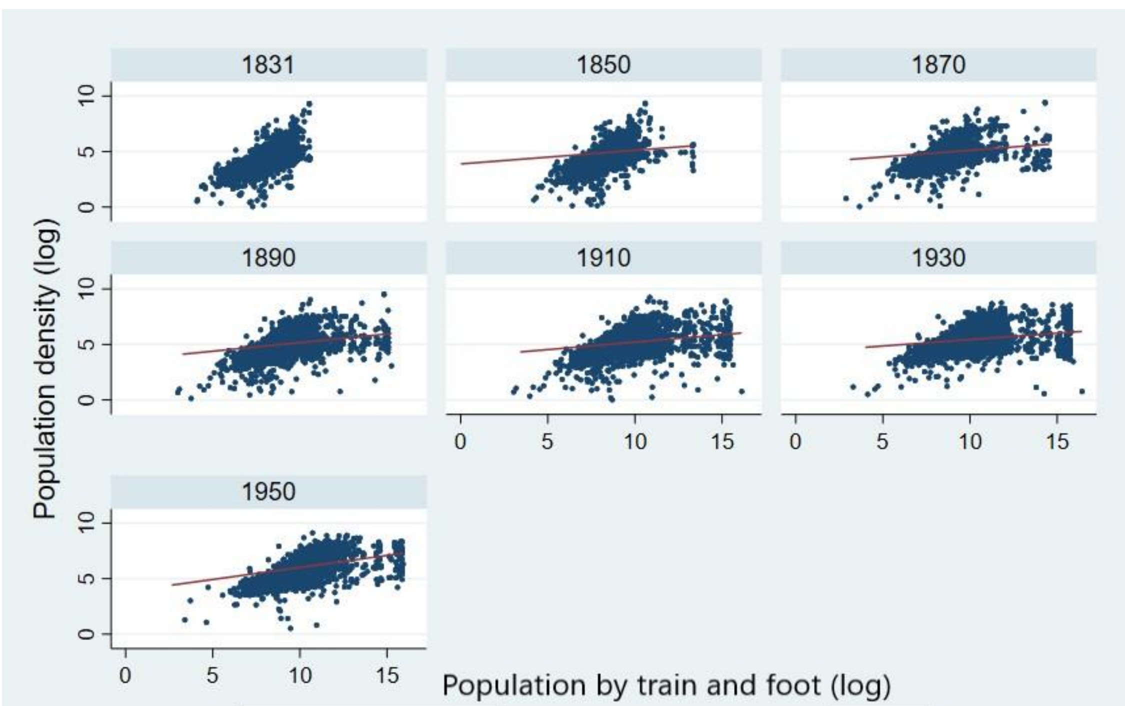


Figure 4.2. Scatterplots by year

Additional Regressions

Year Fixed Effect regressions per province

Table (6) Logarithmic Year Fixed Effect regressions per province, variant of model (2.2)
Dependent: Population per KM2 (log)

	(1) Drenthe	(2) Friesland	(3) Gelderland	(4) Groningen	(5) Limburg	(6) Noord-Brabant	(7) Noord-Holland	(8) Overijssel	(9) Utrecht	(10) Zeeland	(11) Zuid-Holland
Population reachable by train and foot (log)	0.0876***	0.0351*	0.0188**	0.0473***	0.0309***	0.0177	0.113***	0.0366**	0.0522***	0.0172	0.0722***
	(0.0262)	(0.0194)	(0.00837)	(0.0119)	(0.0113)	(0.0182)	(0.0133)	(0.0149)	(0.0154)	(0.0137)	(0.0173)
Population in 45-min walking (log)	0.679***	0.765***	0.686***	0.623***	0.832***	0.727***	0.600***	0.716***	0.692***	0.695***	0.631***
	(0.0310)	(0.0235)	(0.0185)	(0.0190)	(0.0197)	(0.0237)	(0.0260)	(0.0244)	(0.0345)	(0.0240)	(0.0274)
Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,796	3,268	3,269	1,533	1,934	3,026	2,469	1,872	1,188	1,065	2,881
R-squared	0.590	0.596	0.390	0.589	0.682	0.494	0.405	0.446	0.341	0.570	0.406

Standard errors in parentheses, clustered on municipality level
*** p<0.01, ** p<0.05, * p<0.1

Province Fixed Effect regressions per year

Table (8) Logarithmic Province Fixed Effect regressions per year, variant of model (2.2)
Dependent: Population per KM2 (log)

	(1) 1831	(2) 1850	(3) 1870	(4) 1890	(5) 1910	(6) 1930	(7) 1950
Population reachable by train and foot (log)	0.718***	0.00552	0.0292**	0.0529***	0.0402***	0.0275***	0.100***
	(0.0147)	(0.0394)	(0.0137)	(0.0113)	(0.00872)	(0.00716)	(0.0106)
Population in 45-min walking (log)	-[1]	0.705***	0.670***	0.682***	0.726***	0.761***	0.676***
	(-)[1]	(0.0427)	(0.0208)	(0.0192)	(0.0170)	(0.0149)	(0.0176)
Province Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	3,449	3,456	3,477	3,470	3,470	3,482	3,497
R-squared	0.410	0.399	0.420	0.456	0.493	0.555	0.587

Standard errors in parentheses, clustered on municipality level
*** p<0.01, ** p<0.05, * p<0.1

[1]Omitted due to multicollinearity

Year and PC4 Fixed Effect regressions per province

Table (7) Logarithmic Year Fixed Effect regressions per province, variant of model (3)

Dependent: Population per KM2 (log)

	(1) Drenthe	(2) Friesland	(3) Gelderland	(4) Groningen	(5) Limburg	(6) Noord-Brabant	(7) Noord-Holland	(8) Overijssel	(9) Utrecht	(10) Zeeland	(11) Zuid-Holland
Population reachable by train and foot (log)	-0.0109	0.00124	-0.00409	0.00132	0.00115	-0.0251	0.0157	-0.000411	0.00362	0.00159	0.0682***
	(0.00928)	(0.00637)	(0.00377)	(0.00408)	(0.00887)	(0.0196)	(0.0212)	(0.00486)	(0.00507)	(0.00277)	(0.0250)
Population in 45-min walking (log)	1.016***	0.971***	1.074***	1.013***	0.989***	1.022***	0.932***	1.063***	1.012***	1.048***	0.658***
	(0.0212)	(0.0707)	(0.0493)	(0.0266)	(0.116)	(0.0863)	(0.0867)	(0.0913)	(0.0392)	(0.0232)	(0.114)
PC4 Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,796	3,262	3,268	1,533	1,934	3,019	2,467	1,872	1,188	1,064	2,881
R-squared	0.991	0.981	0.946	0.990	0.938	0.906	0.924	0.974	0.964	0.965	0.864

Robust standard errors in parentheses, clustered on municipality level

*** p<0.01, ** p<0.05, * p<0.1